The PHILOSOPHICAL MAGAZINE

FIRST PUBLISHED IN 1798

VOL. 41 SEVENTH SERIES No. 312 JANUARY, 1950

A Journal of Theoretical Experimental and Applied Physics

PRICE 10s.

Annual Subscription £5 2s. 6d. payable in advance.

PRINTED AND PUBLISHED BY TAYLOR & FRANCIS LTD., RED LION COURT, FLEET STREET, LONDON E.C.4

Annals of Science

A QUARTERLY REVIEW OF THE HISTORY OF SCIENCE SINCE THE RENAISSANCE

Contents of Vol. 6, No. 2, July 1949

SOME LETTERS FROM JAKOB SAMUEL WYTTENBACH TO SIR JAMES EDWARD SMITH. By Prof. G. R. de BEER, M.A., D.Sc., F.R.S., F.L.S., F.Z.S.

THOMAS THOMSON (1773–1852). By Prof. J. R. PARTINGTON, M.B.E., D.Sc.

SIR JOHN PRINGLE AND HIS CIRCLE.—PART I. LIFE. By DOROTHEA WALEY SINGER

THE INVENTION OF THE HYGROSCOPE. By F. SHERWOOD TAYLOR, Ph.D., M.A., B.Sc.

A BIOGRAPHICAL NOTE ON WILLIAM BROWNRIGG, M.D., F.R.S. (1711–1800). By J. RUSSELL-WOOD, M.Sc., Ph.D.

THE WORK OF G. T. FECHNER ON THE GALVANIC CIRCUIT. By H. J. J. WINTER, M.Sc., Ph.D.

NOTES

REVIEWS

TAYLOR & FRANCIS, LTD., Red Lion Court, Fleet Street, LONDON, E.C.4

THE

PHILOSOPHICAL MAGAZINE

A JOURNAL OF THEORETICAL EXPERIMENTAL AND APPLIED PHYSICS

First published in 1798

[SEVENTH SERIES—VOL. 41]

I. *Some Conformal Transformations involving Elliptic Functions.*

By Sir Charles Darwin, F.R.S.,
National Physical Laboratory*.

[Received September 19, 1949.]

Abstract.

Conformal transformations are given which are suitable for the discussion of the space outside a *finned rectangle*. These include figures such as rectangle, cross, three-sided box, form like the letter H, "windmill", and the transformation could be applied to any rectangle with four fins of any lengths at its corners, each of them in either of two directions. The transformations convert the outer space into the inside of a rectangle, which can then be transformed back into the outside of a circle, with the same point at infinity as the original space.

By a generalization of a special case of this transformation another is given, whereby an irregular star, composed of any number of rays of any lengths and at any angles branching out from a centre, can be transformed into a circle. This includes, as a simple example, a "dihedral" of two equal arms at any angle.

---

1. There are many uses to which conformal transformations are put in physical problems, and the discovery of the appropriate transformations is not always easy. The only general systematic method of discovering them is the Schwartz Christoffel method, and this is limited to cases

* Communicated by the Author.

where only half the plane is of interest, and to cases where this half-plane can by symmetry be extended to describe the whole plane. As soon as less symmetrical shapes are considered, the discovery of the appropriate transformations becomes guess-work, and it may therefore be useful to put on record a set of such transformations representing a rather wide range of forms.

The present note started with some work of V. M. Falkner* in connection with finned aeroplanes, for which I provided him with certain transformations involving elliptic functions. For this work the Schwartzian method sufficed, but I have had the curiosity to extend it further into a range of shapes for which that method cannot be used. It is hardly necessary to say that, though it was aerodynamics that gave rise to these problems, they are not to be interpreted as giving directly the air flow over an aeroplane of the shape described.

Elliptic functions will rationalize four square roots, and therefore any form involving four right-angled turns can be represented by their means. There is, however, one important further limitation on the shapes here considered, in that I shall only consider the transformations mapping the area *outside* the form and extending to infinity in all directions.

Fig. 1.

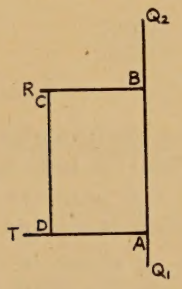
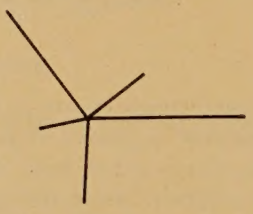


Fig. 2.



The present work gives the transformations whereby the area outside any *finned rectangle* like that in fig. 1 can be transformed into the inside of a rectangle. This rectangle can then in turn be transformed into the area outside a circle in such a manner that the point at infinity coincides with the original point at infinity. In the course of this work the consideration of one special case, the cross, suggested another generalization by which it is possible to transform into a circle a figure that may be described as an *irregular star*, composed of any number of rays of any lengths and at any angles, such as shown in fig. 2. Perhaps the most useful example of this would be an " aeroplane with dihedral angle ".

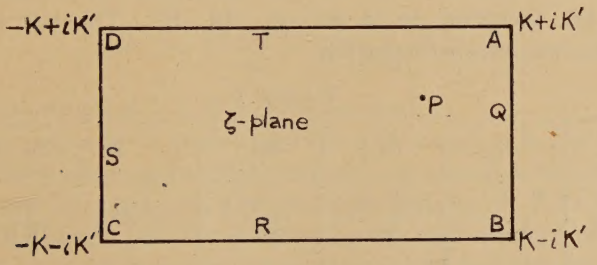
2. Since the purpose of this note is chiefly to put the transformations on record, there is no need to explain in detail how they were found. It need only be said that there are two primitive forms from which they were

* *Aer. Res. Co., Rep. & Mem.* No. 2280 (Oct. 1945).

generalized which may be called the H and the X, from the shapes of those letters. The H was the subject of the previous note, though the notation has been improved from that used there. The interesting case is where the cross-bar is not exactly half way up it. By bisecting the H vertically the Schwartzian method can be used, and the most important point that emerges is that if it is to fit together properly, the elliptic integrals of the third kind which arise in the integration must vanish. This is then seen to be a general principle, because such an integral has a logarithmic singularity, whereas the only admissible singularity is a simple pole to represent the point at infinity. This condition very much simplifies all the work. Other problems involving four right-angled turns can also be solved by elliptic integrals, but in general they involve integrals of the third kind.

The X, or cross, does not lend itself directly to the Schwartzian method, though a symmetrical cross can be discussed in that way by twisting it through 45°. However, as will appear, it is very easy to guess the transformation directly, and there is therefore no need to explain this process.

Fig. 3.



3. In order to follow the various diagrams it will be well to explain their lettering. The four right-angled turns are lettered ABCD taken counter-clockwise in the z-plane. Since the outer region in z is to be transformed into the inside of a rectangle in ζ, these points go clockwise in ζ as shown in fig. 3. The corners of the rectangle are the points $\pm K \pm iK'$ in the ζ-plane. Any point on AB is labelled Q, and similarly for the other sides by R, S and T. In particular these letters will denote the turning points of the fins; if there are more than one on a side of the rectangle, they are given subscripts, and the subscript 0 will be used in the particular case where the mid-point of the side is referred to. One other place of importance in the ζ-rectangle is the pole P, which corresponds in the z-plane to the point at infinity.

4. The transformations all involve expressing z in terms of elliptic functions of ζ. In addition to the ordinary functions, we shall have elliptic integrals of the second kind, and the most convenient one to use is the one

that vanishes at both 0 and K. In the ordinary notation this is written with a term in E/K, but in the present work E will never occur by itself, so we shall adopt a single symbol and write

$$G=E/K. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.1)$$

Then take

$$Z(\zeta)=\int_0^\zeta \mathrm{dn}^2\, u\, du - G\zeta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.2)$$

This function, like $\mathrm{sn}\,\zeta$ etc., is infinite at the points $\pm iK'$, and this is inconvenient. It is quite easy to deal with the infinity of $\mathrm{sn}\,\zeta$, etc., but it is better to eliminate it for Z by considering the function

$$Z(\zeta)+\frac{\mathrm{dn}\,\zeta\,\mathrm{cn}\,\zeta}{\mathrm{sn}\,\zeta} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.3)$$

instead. The second term destroys the infinities at $\pm iK'$, but makes a pole at the origin. It is then comparatively easy to shift this pole to any other point P.

To work out the values in z corresponding to the rectangle in ζ it is convenient to have a table, for though the values of the functions are all given by well known formulæ, there are a good many such formulæ, and their signs are rather confusing. Table I. gives the expressions used; the modulus is taken as k for sn, cn, dn, Z, but as k', its complement, when they are accented, as in sn′, cn′, dn′, Z′. The last row gives the ordinary Jacobian transformation.

TABLE I.

	ζ	$\mathrm{sn}\,\zeta$	$\mathrm{cn}\,\zeta$	$\mathrm{dn}\,\zeta$	$Z(\zeta)+\mathrm{dn}\,\zeta\,\mathrm{cn}\,\zeta/\mathrm{sn}\,\zeta$
Q	$K+i\eta$	$\frac{1}{\mathrm{dn}'\eta}$	$-\frac{ik'\,\mathrm{sn}'\eta}{\mathrm{dn}'\eta}$	$\frac{k'\,\mathrm{cn}'\eta}{\mathrm{dn}'\eta}$	$-i\left\{Z'(\eta)+\frac{\eta\pi}{2KK'}\right\}$
R	$-iK'+\xi$	$\frac{1}{k\,\mathrm{sn}\,\xi}$	$\frac{i\,\mathrm{dn}\,\xi}{k\,\mathrm{sn}\,\xi}$	$\frac{i\,\mathrm{cn}\,\xi}{\mathrm{sn}\,\xi}$	$Z(\xi)+\frac{i\pi}{2K}$
S	$-K+i\eta$	$-\frac{1}{\mathrm{dn}'\eta}$	$\frac{ik'\,\mathrm{sn}'\eta}{\mathrm{dn}'\eta}$	$\frac{k'\,\mathrm{cn}'\eta}{\mathrm{dn}'\eta}$	$-i\left\{Z'(\eta)+\frac{\eta\pi}{2KK'}\right\}$
T	$iK'+\xi$	$\frac{1}{k\,\mathrm{sn}\,\xi}$	$-\frac{i\,\mathrm{dn}\,\xi}{k\,\mathrm{sn}\,\xi}$	$-\frac{i\,\mathrm{cn}\,\xi}{\mathrm{sn}\,\xi}$	$Z(\xi)-\frac{i\pi}{2K}$
	$i\tau$	$\frac{i\,\mathrm{sn}'\tau}{\mathrm{cn}'\tau}$	$\frac{1}{\mathrm{cn}'\tau}$	$\frac{\mathrm{dn}'\tau}{\mathrm{cn}'\tau}$	$-i\left\{Z'(\tau)+\frac{\mathrm{dn}'\tau\,\mathrm{cn}'\tau}{\mathrm{sn}'\tau}+\frac{\pi\tau}{2KK'}\right\}$

The following points may be noted. On the boundary $\mathrm{sn}\,\zeta$ is everywhere real, being always greater than 1 and positive on the right-hand side, negative on the left. $\mathrm{cn}\,\zeta$ is everywhere purely imaginary on the boundary, and its real part is everywhere positive inside. $\mathrm{dn}\,\zeta$ is real on Q and S, imaginary on R and T, while its real part is everywhere positive inside.

The function in the last column is imaginary on Q and S, vanishing at the mid-points, while on R and T it has constant imaginary parts equal and opposite in sign.

5. The transformation yielding the symmetrical cross is

$$z=\frac{\mathrm{dn}\,\zeta}{\mathrm{sn}\,\zeta}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.1)$$

This is easily generalized to an irregular cross as

$$z=\frac{\mathrm{dn}\,\zeta}{\mathrm{sn}\,\zeta-f-ig\,\mathrm{cn}\,\zeta}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.2)$$

where f and g are real, since the denominator is then real all round the contour. For the pole to fall inside the rectangle it is necessary that

$$f^2+g^2<1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.3)$$

The transformation yielding the H with unequal ends is

$$z=\mathrm{Z}(\zeta)+\frac{\mathrm{dn}\,\zeta\,\mathrm{cn}\,\zeta}{\mathrm{sn}\,\zeta-f} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.4)$$

A corresponding generalization for H can be made, and combining the two formulae the final result is

$$\frac{z}{l}=\mathrm{Z}(\zeta)+\frac{\mathrm{dn}\,\zeta\,(\mathrm{cn}\zeta+ig\,\mathrm{sn}\,\zeta+ih)}{\mathrm{sn}\,\zeta-f-ig\,\mathrm{cn}\,\zeta}+m\zeta, \quad . \quad . \quad . \quad (5.5)$$

where all the parameters l, k, f, g, h, m are real.

The first two terms are a generalization of those given in the last column of Table I. By adding and subtracting $\mathrm{dn}\,\zeta\ \mathrm{cn}\,\zeta/\ \mathrm{sn}\,\zeta$ in (5.5) it is easy to see that the second term still destroys the infinity of Z at R_0 and T_0, and now yields a single pole determined by the denominator of the second term. The general formulae corresponding to the last column of Table I. are not written out here, since they are rather complicated, though they may be worked out easily by the use of Table I.

The equation (5.5) has six parameters l, k, f, g, h and m, and by suitable choices these can be adapted to give the six dimensions of the finned rectangle, that is, its two sides and the lengths (and directions) of the four fins.

The turning points of the fins are given by equating to zero

$$\frac{1}{l}\frac{dz}{d\zeta}=m-\mathrm{G}+\{-(1-f^2)+\mathrm{sn}^2\zeta(1-k^2f^2+k'^2g^2)+\mathrm{sn}\,\zeta k'^2(-f+gh)$$
$$+i\,\mathrm{cn}\,\zeta(-h+fg)+i\,\mathrm{sn}\,\zeta\,\mathrm{cn}\,\zeta[-g(1+k'^2)+fhk^2]\}/(\mathrm{sn}\,\zeta-f-ig\,\mathrm{cn}\,\zeta)^2,$$
$$. \quad . \quad . \quad (5.6)$$

which is a quartic in $\mathrm{sn}\,\zeta$.

Writing $p=\sqrt{(1-f^2-g^2)}$ the pole P is at ζ_0, where

$$\mathrm{sn}\,\zeta_0=\frac{f+igp}{1-g^2}, \quad \mathrm{cn}\,\zeta_0=\frac{p-ifg}{1-g^2}. \quad . \quad . \quad . \quad . \quad . \quad (5.7)$$

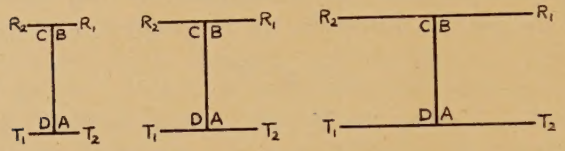
6. We now describe some of the shapes given by specializing the parameters. The diagrams are only sketches, in that when it is asserted that one parameter is changing no attention has been given to keeping the others strictly constant. To draw them accurately would involve much work and give little further information.

Take first the shapes that are symmetrical about two central lines at right angles. For these f, g and h vanish.

The first to be considered is

$$z=\mathrm{Z}(\zeta)+\frac{\mathrm{dn}\,\zeta\,\mathrm{cn}\,\zeta}{\mathrm{sn}\,\zeta}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.1)$$

Fig. 4



As k increases, this gives the sequence of fig. 4. The turning points are the solutions of

$$\frac{\mathrm{cn}^2\zeta}{\mathrm{sn}^2\zeta}=-\mathrm{G} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.2)$$

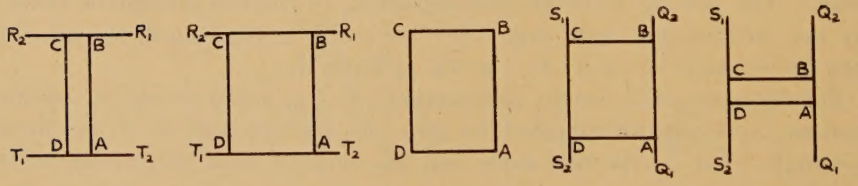
or transforming to the lines R or T by Table I.,

$$\mathrm{dn}^2\,\xi=\mathrm{G} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.3)$$

With this value of ξ,

$$z=\pm\mathrm{Z}(\xi)\pm\frac{i\pi}{2\mathrm{K}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.4)$$

Fig. 5.



at the ends of the fins. When k is not large ξ is somewhere near $\frac{1}{2}\mathrm{K}$, but tables of elliptic functions must be used to evaluate it accurately. The value of k must be very close indeed to 1 to get an H with long arms and a short cross-bar.

Next take

$$z=\mathrm{Z}(\zeta)+\frac{\mathrm{dn}\,\zeta\,\mathrm{cn}\,\zeta}{\mathrm{sn}\,\zeta}+m\zeta \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.5)$$

and consider how it behaves as m changes. The effect of m is to thicken up the rectangle as in fig. 5. It must be positive, for a negative m would cause AB to overlap CD the wrong way round.

The rectangle has no fins when $m=G-k'^2$ after which the fins switch through a right angle. At $m=\pi/2KK'$, BC coincides with AD, and the sequence stops. The second half of the sequence can be directly derived from the first by the Jacobian transformation.

For the special case $k=1/\sqrt{2}$, $m=G-\frac{1}{2}=\pi/4K^2$ the shape is a square, and the Jacobian transformation shows its tetragonal symmetry.

7. If now a term in h is included, the two lines of symmetry are destroyed, but the centre is still a centre of symmetry.

First take

$$z=Z(\zeta)+\frac{\mathrm{dn}\,\zeta\,\mathrm{cn}\,\zeta}{\mathrm{sn}\,\zeta}+\frac{ih\,\mathrm{dn}\,\zeta}{\mathrm{sn}\,\zeta}. \quad . \quad . \quad . \quad . \quad . \quad (7.1)$$

As h increases from zero the fins alter as shown in fig. 6. When $h=(G-k'^2)/kk'$ one pair of fins disappears. As h grows to infinity the side-arms close down, leading to a pure cross when h becomes infinite. If a term in m is added, this merely thickens up the rectangle.

Fig. 6.

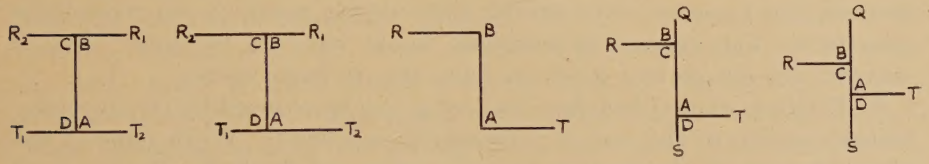
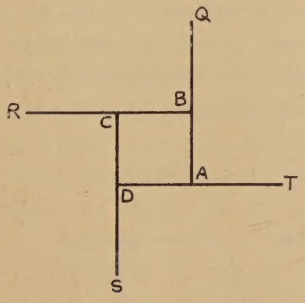


Fig. 7.



Attention may be drawn to the " windmill ", fig. 7:

$$z=Z(\zeta)+\frac{\mathrm{dn}\,\zeta\,\mathrm{cn}\,\zeta}{\mathrm{sn}\,\zeta}+\frac{\pi}{4K^2}\zeta+\frac{ih\,\mathrm{dn}\,\zeta}{\mathrm{sn}\,\zeta}\left(k=\frac{1}{\sqrt{2}}\right),$$

which is tetragonally symmetrical. For those familiar with the peculiarities of the hydrodynamic fluid it is hardly necessary to say that no matter how strongly the wind blows, or in what direction, the windmill experiences no resultant couple at all !

8. Next consider cases where P is no longer at the origin. Take

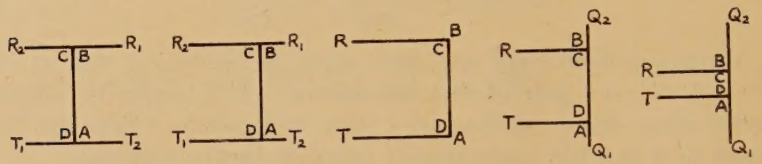
$$z=Z(\zeta)+\frac{\mathrm{dn}\,\zeta\,\mathrm{cn}\,\zeta}{\mathrm{sn}\,\zeta-f}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.1)$$

For growing f the family of shapes is then as in fig. 8. The transitional case with only two fins occurs for

$$f=(G-k'^2)/kG. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.2)$$

If a term in m is added, the rectangle is merely thickened as before. In both these cases a quadratic in $\mathrm{sn}\,\zeta$ determines the values of ξ or η at the turning points, the two pairs of fins being associated with the two signs of $i\,\mathrm{cn}\,\zeta$.

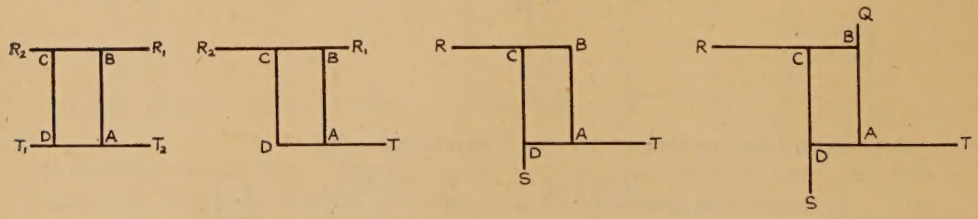
Fig. 8.



If further a term in h is added, the matter becomes more complicated because the equation determining the turning points is an irresoluble quartic, so that numerical solutions would have to be done. Fig. 9 sketches the shapes as h increases from zero to large values.

9. The case where f vanishes, but not g, can be reduced by the Jacobian transformation to the case of g vanishing and not f. When none of the constants vanish the matter becomes very complicated. If numerical values are assigned to all the constants, it is comparatively easy to find the shape. In doing so it is easiest not to use (5.6) in order to find the turning points, but instead merely to estimate the maxima from (5.5) by taking out a few values in their neighbourhoods. Thus fig. 1 is a somewhat exaggerated sketch given by the values

Fig. 9.



$$l=1, \quad k=0{\cdot}5, \quad f=0{\cdot}4, \quad g=0{\cdot}3, \quad h=0{\cdot}1, \quad m=0{\cdot}2.$$

The corners are at $\pm 0{\cdot}337 \pm i0{\cdot}501$ and the fins at

$$Q_1=0{\cdot}337-i\,0{\cdot}565 \qquad R=-0{\cdot}339+i\,0{\cdot}501$$

$$Q_2=0{\cdot}337+i\,0{\cdot}866 \qquad T=-0{\cdot}386-i\,0{\cdot}501$$

The converse problem of finding the parameters for a given shape is much more troublesome. I explored the method of solution by taking

the comparatively simple problem of finding the parameters for a cross with arms of lengths 1, 2, 3, 4. This has the transformation

$$z = \frac{ih\,\mathrm{dn}\,\zeta}{\mathrm{sn}\,\zeta - f - ig\,\mathrm{cn}\,\zeta}.$$

Starting values were guessed for k, f, g by considering the values for the half-symmetrical cases, where one pair of opposite arms are equal while the other pair have the required ratio, and with these the arms were calculated. It was also possible to set down an equation for the variation of the parameters, so that the effect of δf, etc., on all four arms could be seen; most of the quantities for these variations occur already in the main calculation. Three stages led to the values

$$h = 2{\cdot}835, \quad k = 0{\cdot}853, \quad f = 0{\cdot}489, \quad g = 0{\cdot}294,$$

which are correct to about one figure in the last place. The ends of the arms are then at 4, $3i$, -2, $-i$.

A similar method should work in the general case. The corners of the rectangle immediately give two relations between l, m and k. For the rest the method of trial must be used, but there is one control on the starting values of f, etc., which would help considerably. The value of $dz/d\zeta$ takes a fairly simple form at each of the corners, by the consideration that in differentiating (5.5) $\mathrm{dn}\,\zeta$ vanishes at the corners. Then according as $dz/d\zeta$ changes sign from one corner to the next or not, there will be one fin on that side or else either two or none. In this way some guidance can be given as to appropriate starting values for k, f, g, h, and the variation of parameters can then be used to get improvement. It may be noted that variation of k is a troublesome process, even if $\mathrm{sn}\,\zeta$ instead of ζ is used as the independent variable, so that for this one parameter it might be simplest in practice to omit this variation, and merely to find the shape twice over for two different close values of k. However, there is no doubt that the solution would be arduous, and it seemed hardly worth while to make up a purely artificial problem in order to solve it.

10. It will often be convenient to apply a second transformation, converting the ζ-rectangle into the exterior of a circle in such a manner that the point at infinity becomes the old point at infinity of the z-plane.

A simple case will show the method. The symmetrical cross is given by

$$z = \frac{\mathrm{dn}\,\zeta}{\mathrm{sn}\,\zeta}. \qquad \ldots\ldots\ldots \quad (10.1)$$

Write

$$z_0 = \frac{1 + \mathrm{cn}\,\zeta}{\mathrm{sn}\,\zeta}. \qquad \ldots\ldots\ldots \quad (10.2)$$

Then the point at infinity in z_0 is the same as in z. On the sides of the rectangle $\mathrm{sn}\,\zeta$ is real and $\mathrm{cn}\,\zeta$ imaginary, so that $|z_0|^2 = 1$, and so the rectangle becomes a circle of unit radius in the z_0-plane.

In the general case the transformation is

$$z_0=\frac{1}{\sqrt{(1-g^2)}}\frac{(1-g^2)-(f-igp)\operatorname{sn}\zeta+(p+ifg)\operatorname{cn}\zeta}{\operatorname{sn}\zeta-f-ig\operatorname{cn}\zeta}, \quad . \quad (10.3)$$

where $p=\sqrt{(1-f^2-g^2)}$.

11. If it is desired to expose the finned rectangle to a "wind", it is necessary to form a function that has a pole at P, and that is real on the rectangle. Such a function is

$$w=\frac{\mathrm{A}\operatorname{sn}\zeta+i\mathrm{B}\operatorname{cn}\zeta}{\operatorname{sn}\zeta-f-ig\operatorname{cn}\zeta}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (11.1)$$

in which A and B are real. From this the components of the wind at infinity are given by the real and imaginary parts of

$$\left(\frac{dw}{dz}\right)_\infty=\frac{\mathrm{A}\operatorname{sn}\zeta_0+i\mathrm{B}\operatorname{cn}\zeta_0}{\operatorname{dn}\zeta_0(\operatorname{cn}\zeta_0+ig\operatorname{sn}\zeta_0+ih)}, \quad . \quad . \quad . \quad . \quad (11.2)$$

where ζ_0 is given by (5.7), $\operatorname{dn}\zeta_0$ having its real part positive.

An irrotational circulation round the rectangle is given by

$$w=i\mathrm{C}\ln z_0,$$

where z_0 is given by (10.3), since $\ln z_0$ is pure imaginary on the rectangle.

12. If it is desired to transform the finned rectangle into a circle, the intermediate use of the rectangle is in general indispensable, but there is an exception in the case of the cross, since the transformation then involves no elliptic integrals of the second kind. For example, between (10.1) and (10.2) ζ may be eliminated. Taking the special case of the equal armed cross with $k=1/\sqrt{2}$, the eliminant is

$$z=\tfrac{1}{2}(z_0^2+z_0^{-2})^{\frac{1}{2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.1)$$

This may be compared with $z=\frac{1}{2}(z_0+z_0^{-1})$, which transforms the circle into a line bounded at both ends. This suggests immediately a generalization to a star of six, eight, etc., equal rays, but a very much wider generalization than this can be made.

Take

$$z=lz_0^{-1}(z_0-e^{i\alpha_1})^{\lambda_1}(z_0-e^{i\alpha_2})^{\lambda_2}\ldots(z_0-e^{i\alpha_n})^{\lambda_n}. \quad . \quad . \quad . \quad (12.2)$$

where

$$\lambda_1+\lambda_2+\ldots+\lambda_n=2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (12.3)$$

To study the transformation put $z=e^{i\theta}$, and consider the part where $\alpha_1<\theta<\alpha_2$. For this part, using (12.3),

$$z=4l\exp\{i\tfrac{1}{2}\Sigma_1^n\alpha_s\lambda_s+i\pi(\lambda_1-1)\}\sin^{\lambda_1}\tfrac{1}{2}(\theta-\alpha_1)\sin^{\lambda_2}\tfrac{1}{2}(\alpha_2-\theta)\ldots\sin^{\lambda_n}\tfrac{1}{2}(\alpha_n-\theta),$$
$$. \quad . \quad . \quad (12.4)$$

so that z traces a ray in the direction $\frac{1}{2}\Sigma_1^n\alpha_s\lambda_s+\pi(\lambda_1-1)$, vanishing at $\theta=\alpha_1$ and α_2, and so having a maximum in between. When $\alpha_2<\theta<\alpha_3$ there will be another ray in the direction $\frac{1}{2}\Sigma_1^n\alpha_s\lambda_s+\pi(\lambda_2+\lambda_1-1)$. Thus

there are rays of various lengths at successive angles $\pi\lambda_1$, $\pi\lambda_2$, . . . $\pi\lambda_n$ to one another. By suitable choice of l, $n-1$ of the λ's and the n α's any arbitrary irregular star of the type shown in fig. 2 can be transformed into a circle.

In finding this transformation the λ's can be given at once, but the α's would have to be found by trial. It may be noted that if two adjacent α's are taken equal, one of the rays disappears, and so it may be concluded that if two α's are close together, the ray between them will be short. We will be content with two examples.

Take

$$z=lz_0^{-1}(z_0-1)^{2\beta}(z_0+1)^{2-2\beta}. \quad . \quad . \quad . \quad . \quad . \quad (12.5)$$

This represents a dihedral with equal arms at angle $2\pi\beta$ apart, as in fig. 10. The maxima occur at $z_0=\exp(\pm i\theta_1)$, where $\cos\theta_1=1-2\beta$. The rays are of length $4l\beta^{\beta}(1-\beta)^{1-\beta}$.

Fig. 10.

Fig. 11.

As a second example I arbitrarily chose the problem of an arrow-head, in which the barbs, at 30° to the shaft, were only half as long, as in fig. 11. This is represented by the transformation

$$z=z_0^{-1}(z_0-e^{i\alpha})^{1/6}(z_0+1)^{5/3}(z_0-e^{-i\alpha})^{1/6}, \quad . \quad . \quad . \quad (12.6)$$

and without much trouble it was found that $\alpha=71°\,46'$. The shaft is then of length 3·348 and the barbs 1·673.

II. *The Propagation of Sound Waves in an Open-ended Channel.*

By W. Chester,
Department of Mathematics, Leeds University *.

[Received August 22, 1949.]

Summary.

An exact solution is obtained for the problem of a harmonic wave of sound approaching the open end of a two-dimensional channel, and suffering reflection and transmission. Within a certain wavelength

* Communicated by Professor S. Goldstein, F.R.S.

range, the wave returning along the channel becomes sensibly plane at large distances from the mouth, the reflection coefficient being $\exp[-2\pi b/\lambda]$, where $2b$ is the width of the channel and λ is the wavelength. Explicit formulæ are also obtained for the change of phase in the returning wave, and for the transmitted energy.

It is also shown that the solution can be used to deduce the energy absorbed from a plane wave external to the channel and approaching from an arbitrary direction.

Introduction.

When a plane harmonic wave of sound is incident upon a two-dimensional channel, formed by two semi-infinite parallel planes, diffraction effects will occur at the open end. Ultimately, a periodically steady state is reached, and the object of the present paper is to determine the asymptotic form of the resulting periodically steady potential field, assuming that the wavelength lies within a certain range so that only the dominant mode is propagated inside the channel. There is no theoretical difficulty in extending the analysis to include a wider wavelength range, but, from a physical view-point, the present range is considered sufficient.

Problems of this type were first considered by Rayleigh using approximate methods. For the particular case of an incident wave originating inside a circular pipe, Rayleigh (1940) gave an estimate of the end correction when the wavelength was large, and also deduced that it increased indefinitely with the wavelength for a two-dimensional channel (Rayleigh 1904). (The end correction is the correction to the length of the pipe in order that the open end may be treated as a loop).

More recently, Levine and Schwinger (1948) gave an exact solution for the circular pipe, valid for a wave-length range in which only dominant mode propagation occurs. The present investigation gives the analogous solution for the two-dimensional channel, and is inspired by the analysis of the above authors.

A detailed investigation is given for an incident wave approaching the open end from inside the channel. The whole of the asymptotic behaviour of the wave, both inside and outside the channel, can be described in terms of the discontinuity in potential across the channel walls, and the problem consists, essentially, of deriving an integral equation for this quantity which is susceptible to rigorous, analytical solution. Rayleigh's conclusion is verified and, in general, the final results assume a simple form. The end correction and transmitted energy are exhibited graphically in figures 5 and 6.

It is also shown that, having obtained the solution for the above problem, the energy absorbed by the channel from a plane wave approaching the open end from an arbitrary direction can then be deduced immediately with the aid of a simple reciprocity relation which correlates the radiating and absorbing properties of the channel.

Asymptotic Behaviour.

Since the analysis is eventually confined to the behaviour at large distances from the mouth of the channel, we begin with some general remarks concerning the asymptotic form assumed by the potential both inside and outside the channel.

The origin of coordinates is taken at the mouth of the channel with the negative x-axis along the centre-line. The potential of the incident wave is denoted by $A \exp [ik(x-at)]$ or $\exp [-ik\{r \cos (\theta-\theta_0)-at\}]$ according as it originates inside the channel or approaches from a direction θ_0, and the final potentials resulting from these two incident waves are denoted by ϕ_a, ϕ_b respectively.

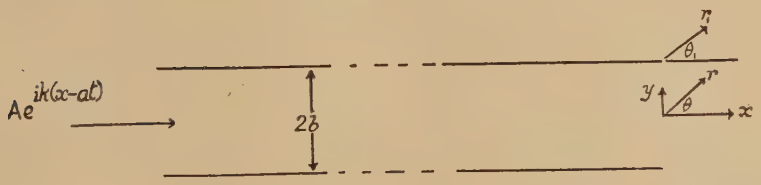
Inside the channel, the most general solution of the wave equation $(\nabla^2+k^2)\phi=0$, expanded as a Fourier series in y and satisfying the boundary condition of zero normal derivative at $y=\pm b$, is of the form

$$\phi=A \exp [ikx]+B \exp [-ikx]+ \sum_{n=1}^{\infty} (A_n \exp [x(q^2-k^2)^{\frac{1}{2}}]$$
$$+B_n \exp [-x(q^2-k^2)^{\frac{1}{2}}] \cos qy+ \sum_{n=0}^{\infty} (A'_n \exp [x(s^2-k^2)^{\frac{1}{2}}]$$
$$+B'_n \exp [-x(s^2-k^2)^{\frac{1}{2}}]) \sin sy, \quad x<0, \qquad \ldots\ldots\ldots \quad (1)$$

where

$$q=n\pi/b, \quad s=(2n+1)\pi/2b.$$

Fig. 1.



(The factor $\exp [ikat]$ here, and in the subsequent analysis, is suppressed).

The potential ϕ_a is symmetrical about the centre line of the channel, so that, in this case, $A'_n=B'_n=0$. It is now clear from equation (1), that the condition for dominant mode propagation is $bk<\pi$. All the coefficients B_n must then be zero if ϕ_a is to remain finite for large negative x, and so we deduce that, as $x \to -\infty$ inside the channel,

$$\phi_a \sim A \exp [ikx]+B \exp [-ikx]. \qquad \ldots\ldots \quad (2)$$

Similar remarks apply to ϕ_b except that A'_n is no longer zero. This reduces the range of k to $kb<\pi/2$; then

$$\phi_b \sim C \exp [-ikx], \qquad \ldots\ldots\ldots \quad (3)$$

since there is now no incident wave inside the channel.

In practice, the above restrictions will usually be satisfied. The condition $bk<\pi/2$, for example, is equivalent to $b/\lambda<\frac{1}{4}$, and for a typical

sound frequency of 270 oscillations per second, the wavelength is four feet.

Outside the channel, ϕ_a will behave, asymptotically, like a source situated at the origin whose amplitude varies with direction. Thus

$$\phi_a \sim f(\theta)\frac{\exp[ikr]}{r^{\frac{1}{2}}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The behaviour of ϕ_b is, however, somewhat more complex. If $\frac{\pi}{2}<\theta_0<\pi$, it must take account of the incident wave which persists in the region $-(\pi-\theta_0)<\theta_1<\pi$, and a wave reflected from the upper wall in the region $(\pi-\theta_0)<\theta_1<\pi$, where θ_1 is the polar coordinate referred to the upper lip of the channel (fig. 1). It follows that the correction term due to diffraction at the mouth of the channel cannot tend to zero at infinity along $\theta_1=\pm(\pi-\theta_0)$ since the continuity of the potential must be maintained in the transition from one region to another.

To overcome this difficulty, appeal is made to the known solution for the diffraction of a harmonic wave by a semi-infinite wall. This problem was first solved by Sommerfeld (1894) and later discussed by Carslaw (1898). In terms of coordinates (r_1, θ_1) about an origin at the edge of the wall (fig. 1), a wave $\exp[-ikr_1\cos(\theta_1-\theta_0)]$ is diffracted to give a potential,

$$\left(\frac{k}{\pi}\right)^{\frac{1}{2}}\exp[-i\pi/4]\left\{\exp[-ikr_1\cos(\theta_1-\theta_0)]\int_{-\infty}^{(2r_1)^{\frac{1}{2}}\cos\frac{1}{2}(\theta_1-\theta_0)}\exp[ik\lambda^2]\,d\lambda\right.$$

$$\left.+\exp[-ikr_1\cos(\theta_1+\theta_0)]\int_{(2r_1)^{\frac{1}{2}}\cos\frac{1}{2}(\theta_1+\theta_0)}^{\infty}\exp[ik\lambda^2]\,d\lambda.\right\} \quad . \quad . \quad (5)$$

It is easily verified that the normal derivative of this expression is zero along the wall $(\theta_1=\pm\pi)$, and since

$$\int_{-\infty}^{(2r_1)^{\frac{1}{2}}\cos\frac{1}{2}\theta}\exp[ik\lambda^2]\,d\lambda \sim \left(\frac{\pi}{k}\right)^{\frac{1}{2}}\exp[i\pi/4]+\frac{\exp[ikr_1(1+\cos\theta)]}{2^{\frac{3}{2}}ik^{\frac{1}{2}}r^{\frac{1}{2}}\cos\frac{1}{2}\theta}\ 0<\theta<\pi-\epsilon,$$

$$\sim \frac{\exp[ikr_1(1+\cos\theta)]}{2^{\frac{3}{2}}ik^{\frac{1}{2}}r^{\frac{1}{2}}\cos\frac{1}{2}\theta}\ \pi+\epsilon<\theta<\pi, \quad . \quad . \quad . \quad (6)$$

it follows that the above potential includes the plane wave contributions in the relevant intervals, together with a term which behaves like $\exp[ikr]/r^{\frac{1}{2}}$ times a function of θ_1, except along the lines $\theta_1=\pm(\pi-\theta_0)$. This last peculiarity of the solution is essential if the potential is to remain continuous in the transition from one region to another.

In the present problem there will be further diffraction effects in addition to those given by (5), due to the presence of the second diffracting wall of the channel. Their contribution to the potential must,

however, take the form $g(\theta_1)\exp[ikr_1]/r^{\frac{1}{2}}$ for large r, and the asymptotic behaviour of ϕ_b is, accordingly,

$$\phi_b \sim \left(\frac{k}{\pi}\right)^{\frac{1}{2}} \exp[-i\pi/4 - ikb\sin\theta_0]\Big\{\exp[-ikr_1\cos(\theta_1-\theta_0)] \times \int_{-\infty}^{(2r_1)^{\frac{1}{2}}\cos\frac{1}{2}(\theta_1-\theta_0)} \exp[ik\lambda^2]\,d\lambda + \exp[-ikr_1\cos(\theta_1+\theta_0) \times \int_{(2r_1)^{\frac{1}{2}}\cos\frac{1}{2}(\theta_1+\theta_0)}^{\infty} \exp[ik\lambda^2]\,d\lambda\Big\} + g(\theta_1)\frac{ikr_1}{r_1^{\frac{1}{2}}} \quad . \quad . \quad . \quad . \quad . \quad (7)$$

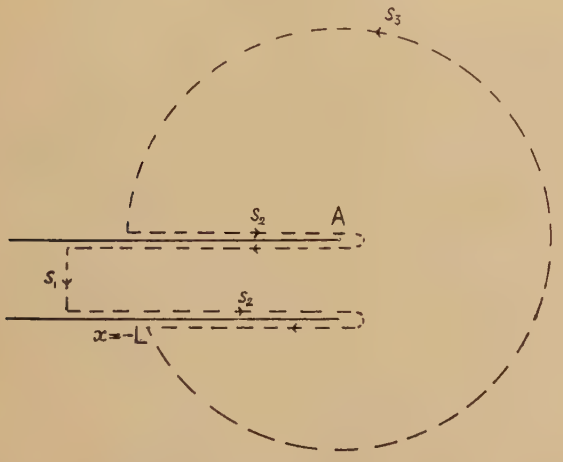
outside the channel.

The factor $\exp[-ikb\sin\theta_0]$ is due to the change in origin. An incident wave $\exp[-ikr\cos(\theta-\theta_0)]$ becomes

$$\exp[-ikr_1\cos(\theta_1-\theta_0)-ikb\sin\theta_0]$$

when referred to coordinates (r_1, θ_1) at the upper lip of the channel.

Fig. 2.



Strictly, equation (7) has been derived for $\pi/2<\theta_0<\pi$ since, for $0<\theta_0<\pi/2$, the boundary for the incident wave term in ϕ_b is a line parallel to $\theta_1=-(\pi-\theta_0)$ and passing through the lower lip of the channel. It can, however, be shown that the above expression does represent the *asymptotic* behaviour of ϕ_b for $0<\theta_0<\pi$.

Reciprocity Relation.

In this section we show that the radiation characteristics of the channel can be used to determine the energy absorbed from a wave originating outside the channel.

To obtain this reciprocity relation, we apply Green's Theorem,

$$\iint_S \{\phi_a\nabla^2\phi_b - \phi_b\nabla^2\phi_a\}\,dS = \int_s \left\{\phi_a\frac{\partial\phi_b}{\partial n} - \phi_b\frac{\partial\phi_a}{\partial n}\right\}ds \quad . \quad . \quad (8)$$

to the two potentials ϕ_a, ϕ_b, the domain of integration being that bounded

by the curves s_1, s_2, s_3 (fig. 2), and n being the outward normal. The portion of the bounding curve denoted by s_2 is taken along the upper and lower sides of each wall of the channel, while s_3 is a large circle, centre A, the origin of the coordinates (r_1, θ_1). Both s_1 and s_3 are considered to be at a sufficiently large distance from the origin to enable the asymptotic relations for ϕ_a, ϕ_b to be used.

Since both ϕ_a and ϕ_b satisfy the equation

$$(\nabla^2+k^2)\phi=0$$

in the region considered, and since, further, they have zero normal derivatives at the walls, equation (8) reduces to

$$\int_{s_1}\left\{\phi_a\frac{\partial\phi_b}{\partial n}-\phi_b\frac{\partial\phi_a}{\partial n}\right\}ds+\int_{s_3}\left\{\phi_a\frac{\partial\phi_b}{\partial n}-\phi_b\frac{\partial\phi_a}{\partial n}\right\}ds=0. \quad . \quad . \quad (9)$$

Substituting from equations (2) and (3), the first integral becomes

$$-\int_{-b}^{b}\Big[(\mathrm{A}\exp[ikx]+\mathrm{B}\exp[-ikx])\frac{\partial}{\partial x}(\mathrm{C}\exp[-ikx])-\mathrm{C}\exp[-ikx]$$
$$\times\frac{\partial}{\partial x}(\mathrm{A}\exp[ikx]+\mathrm{B}\exp[-ikx])\Big]_{x=-\mathrm{L}}dy=4ibk\mathrm{AC}. \quad . \quad . \quad . \quad (10)$$

In evaluating the second integral in equation (9) we note, from equation (7), that the behaviour of ϕ_b on s_3 is equivalent to that of the terms $\exp[-ikr_1\cos(\theta_1-\theta_0)-ikb\sin\theta_0]$ and $\exp[-ikr_1\cos(\theta_1+\theta_0)-ikb\sin\theta_0]$ in the intervals $-(\pi-\theta_0)+\epsilon<\theta_1<\pi$ and $(\pi-\theta_0)+\epsilon<\theta_1<\pi$ respectively together with a term which behaves like $h(\theta_1)\exp[ikr_1]\,r_1^{-\frac{1}{2}}$, except in the intervals $|\theta_1\pm(\pi-\theta_0)|<\epsilon$. It can be shown that the contribution to the integral along s_3 from the two small intervals $|\theta_1\pm(\pi-\theta_0)|<\epsilon$ is $0(\epsilon)$ in the limit $r_1\to\infty$, while the contribution from the term $h(\theta_1)\exp[ikr_1]/r_1^{\frac{1}{2}}$ is

$$\int\Big\{f(\theta_1)\frac{\exp[ikr_1+ikb\sin\theta_1]}{r_1^{\frac{1}{2}}}\frac{\partial}{\partial r_1}\Big(h(\theta_1)\frac{\exp[ikr_1]}{r_1^{\frac{1}{2}}}\Big)$$
$$-h(\theta_1)\frac{\exp[ikr_1]}{r_1^{\frac{1}{2}}}\frac{\partial}{\partial r_1}\Big(f(\theta_1)\frac{\exp[ikr_1+ikb\sin\theta_1]}{r_1^{\frac{1}{2}}}\Big)\Big\}r_1\,d\theta_1=0, \quad . \quad . \quad (11)$$

integrated over any range of θ_1, since the integrand is identically zero.

Again, a term $\exp[-ikr_1\cos\theta_1]$ (say) in ϕ_b will give a contribution

$$\int\Big\{f(\theta_1)\frac{\exp[ikr_1+ikb\sin\theta_1]}{r_1^{\frac{1}{2}}}\frac{\partial}{\partial r_1}(\exp[-ikr_1\cos\theta_1])-\exp[ikr_1\cos\theta_1]$$
$$\times\frac{\partial}{\partial r_1}\Big(f(\theta_1)\frac{\exp[ikr_1+ikb\sin\theta_1]}{r_1^{\frac{1}{2}}}\Big)\Big\}r_1\,d\theta_1$$
$$=-ikr_1^{\frac{1}{2}}\int f(\theta_1)\exp[ikb\sin\theta_1+ikr_1(1-\cos\theta_1)](1+\cos\theta_1)\,d\theta_1+0(r^{-\frac{1}{2}})$$
$$=2^{\frac{3}{2}}ik\int f(\theta_1)\exp[ikb\sin\theta_1]\cos\theta_1/2\,\frac{\partial}{\partial\theta}\Big[\pm\int_{(2r_1)^{\frac{1}{2}}|\sin\frac{\theta_1}{2}|}^{\infty}\exp[ik\lambda^2]\,d\lambda\Big]$$
$$\times d\theta_1+0(r^{-\frac{1}{2}}), \quad . \quad . \quad (12)$$

the plus or minus sign to be taken according as θ_1 is positive or negative.

A simple integration by parts now shows that this expression tends to zero as $r_1 \to \infty$ unless the range of integration includes the point $\theta_1=0$.

Thus, for $0<\theta_0<\pi$, the only contribution to the integral along s_3, in the limit as $r_1 \to \infty$, occurs from the term

$$\exp\,[-ikr_1 \cos\,(\theta_1-\theta_0)-ikb \sin\,\theta_0],$$

in a small interval surrounding $\theta_1=\theta_0$. The term

$$\exp\,[-ikr_1 \cos\,(\theta_1+\theta_0)-ikb \sin\,\theta_0]$$

has no contribution, in the limit, since it occurs only for $(\pi-\theta_0)<\theta_1<\pi$ which exludes the point $\theta_1=-\theta_0$.

It follows that

$$\lim_{r_1\to\infty}\int_{s_3}\left(\phi_a\frac{\partial\phi_b}{\partial n}-\phi_b\frac{\partial\phi_a}{\partial n}\right)ds$$

$$=-\lim_{r_1\to\infty} ikr_1^{\frac{1}{2}}\int_{\theta_0-\delta}^{\theta_0+\delta} f(\theta_1)\exp\,[ikb(\sin\,\theta_1-\sin\,\theta_0)+ikr_1\{1-\cos\,(\theta_1-\theta_0)\}]$$
$$\times(1+\cos\,(\theta_1-\theta_0))\,d\theta_1$$

$$=-\lim_{r_1\to\infty} 2ikr_1^{\frac{1}{2}}f(\theta_0)\int_{-\delta}^{\delta}\exp\,[2ikr_1\sin^2\theta/2]\cos\theta/2\,d\theta$$

$$=-4ikf(\theta_0)\int_{-\infty}^{\infty}\exp\,[2ikz^2]\,dz$$

$$=-4ik\left(\frac{\pi}{2k}\right)^{\frac{1}{2}}f(\theta_0)\exp\,[i\pi/4]. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (13)$$

Finally, equating (10) and (13), we have

$$C=\left(\frac{\pi}{2k}\right)^{\frac{1}{2}}\frac{\exp\,[i\pi/4]f(\theta_0)}{Ab}. \quad \ldots\ldots\ldots \quad (14)$$

The limiting case $\theta_0=\pi$ is anomalous. There is now no reflected wave, but the term which takes account of this wave in equation (5) does not tend to zero for $\theta_0=\pi$ and the whole expression for ϕ_b is asymptotic to $2\exp\,[ikr_1\cos\theta_1]$ for $0<\theta_0<\pi$. The asymptotic behaviour of ϕ_b for, an incident wave, $\exp\,[ikr_1\cos\theta_1]$ is, in fact,

$$\phi_b\sim\left(\frac{k}{\pi}\right)^{\frac{1}{2}}\exp\,[-i\pi/4+ikr_1\cos\,\theta_1]\int_{-\infty}^{(2r_1)^{\frac{1}{2}}\sin\frac{1}{2}\theta}\exp\,[ik\lambda^2]\,d\lambda+g(\theta_1)\frac{\exp\,[ikr_1]}{r_1^{\frac{1}{2}}}$$
$$\ldots \quad (15)$$

which, for large r, behaves, as it should, like $\exp\,[ikr_1\cos\theta_1]$ for $0<\theta_1<\pi$.

The analogous relation to (14) is then

$$C(\pi)=\left(\frac{\pi}{2k}\right)^{\frac{1}{2}}\frac{\exp\,[i\pi/4]f(\pi)}{2Ab}. \quad \ldots\ldots \quad (16)$$

Equations (14) and (16) enable us to deduce the energy absorbed from

a wave approaching from any direction θ_0, in terms of the asymptotic behaviour of a wave originating inside the channel.

Expressions for the Reflected and Transmitted Waves.

The analysis is now confined to the asymptotic behaviour of the wave originating inside the channel which, as we shall see, can be described wholly in terms of the discontinuity in ϕ_a across the walls of the channel.

To prove this, we apply Green's Theorem to the potential ϕ_a, and a plane wave $\exp[-ikr\cos(\theta-\theta_0)]$. The domain of integration is that shown in fig. 2 except that the centre of the large circle is now taken to be 0, the origin of the (r, θ) coordinates.

The surface integral again vanishes and, using the asymptotic expressions for ϕ_a and the fact that its normal derivative vanishes along the walls of the channel, we find that

$$\int_{s_2}\left[f(\theta)\frac{\exp[ikr]}{r^{\frac{1}{2}}}\frac{\partial}{\partial r}(\exp[-ikr\cos(\theta-\theta_0)])-\exp[-ikr\cos(\theta-\theta_0)]\right.$$
$$\left.\times\frac{\partial}{\partial r}\left(f(\theta)\frac{\exp[ikr]}{r^{\frac{1}{2}}}\right)\right]ds$$
$$-\int_{-\mathrm{L}}^{0}[\phi_a]_{y=b}\frac{\partial}{\partial y}\exp[-ikx\cos\theta_0-iky\sin\theta_0]_{y=b}\,dx$$
$$-\int_{-\mathrm{L}}^{0}[\phi_a]_{y=-b}\frac{\partial}{\partial y}\exp[-ikx\cos\theta_0-iky\sin\theta_0]_{y=-b}\,dx$$
$$-\int_{s_1}\left\{(\mathrm{A}\exp[ikx]+\mathrm{B}\exp[-ikx])\frac{\partial}{\partial x}\exp[-ikx\cos\theta_0-iky\sin\theta_0]\right.$$
$$-\exp[-ikx\cos\theta_0-iky\sin\theta_0]\frac{\partial}{\partial x}(\mathrm{A}\exp[ikx]$$
$$\left.+\mathrm{B}\exp[-ikx])\right\}_{x=-\mathrm{L}}ds=0, \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (17)$$

where

$$[\phi_a]=\phi_a)_{y+0}-\phi_a)_{y-0} \quad \text{and} \quad [\phi_a]_{y=b}=-[\phi_a]_{y=-b}=\mathrm{H}(x) \text{ (say)}. \qquad \ldots \quad (18)$$

The first integral, by a similar argument to that used in the derivation of (13) is, in the limit as $r\to\infty$,

$$-2^{\frac{3}{2}}\pi^{\frac{1}{2}}ik^{\frac{1}{2}}f(\theta_0)\exp[i\pi/4]. \qquad \ldots\ldots\ldots \quad (19)$$

Substituting in equation (17), and simplifying the remaining terms, we get

$$\exp[3i\pi/4](2k\pi)^{\frac{1}{2}}f(\theta)=\lim_{\mathrm{L}\to\infty}\left[k\sin\theta\sin(kb\sin\theta)\int_{-\mathrm{L}}^{0}\mathrm{H}(x)\right.$$
$$\times\exp[-ikx\cos\theta]\,dx+\frac{i\sin(kb\sin\theta)}{\sin\theta}\left\{\mathrm{A}(1+\cos\theta)\exp[-ik\mathrm{L}(1-\cos\theta)]\right.$$
$$\left.\left.-\mathrm{B}(1-\cos\theta)\exp[ik\mathrm{L}(1+\cos\theta)]\right\}\right] \qquad \ldots\ldots\ldots \quad (20)$$
$$=\mathrm{F}(k\cos\theta) \quad \text{(say)}. \qquad \ldots\ldots\ldots \quad (21)$$

Since the asymptotic behaviour of $H(x)$ will be the same as that of ϕ_a inside the channel, reference to equation (2) shows that the oscillating terms on the right-hand side of equation (20) are cancelled by equal and opposite contributions from the integral, and the limit is uniquely defined.

It is convenient, at this point, to assume that k has a small positive imaginary part which will eventually be reduced to zero at an appropriate point in the analysis. Considered as a function of the complex variable ζ, $F(\zeta)$ then becomes, for $Im\zeta > Imk$,

$$F(\zeta)=(k^2-\zeta^2)^{\frac{1}{2}} \sin\{b(k^2-\zeta^2)^{\frac{1}{2}}\}H(\zeta), \quad . \quad . \quad . \quad . \quad (22)$$

where

$$H(\zeta)=\int_{-\infty}^{0} H(x) \exp[-i\zeta x]\, dx, \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

and is the Fourier transform of a function equal to $H(x)$ when x is negative and zero for positive x. For economy of notation the same symbol is used for a function and its transform, the distinction being made by the argument of the function.

We may, therefore, regard (20) as the analytic continuation of $F(\zeta)$ in the interval $-Imk \leqslant Im\zeta \leqslant Imk$, so that, for $\zeta = k\cos\theta$

$$F(k\cos\theta)=\exp[3i\pi/4](2k\pi)^{\frac{1}{2}}f(\theta)=k\sin\theta\sin(kb\sin\theta)H(k\cos\theta). \quad (24)$$

The asymptotic form of $H(x)$ given by equation (2), shows, with the help of (23), that $H(\zeta)$ has simple poles at $\zeta=\pm k$. Further, from (20), (22)

$$\left.\begin{aligned} \exp[3i\pi/4](2k\pi)^{\frac{1}{2}}f(0)&=2ikbA=-2bk\lim_{\zeta\to k}(\zeta-k)H(\zeta)=-2bk\operatorname*{Res}_{\zeta=k}H(\zeta),\\ \exp[3i\pi/4](2k\pi)^{\frac{1}{2}}f(\pi)&=-2ikbB=2bk\lim_{\zeta\to -k}(\zeta+k)H(\zeta)=2bk\operatorname*{Res}_{\zeta=-k}H(\zeta), \end{aligned}\right\} \quad . \quad . \quad . \quad (25)$$

and the reflection coefficient, R, becomes

$$R=\frac{B}{A}=\frac{\operatorname*{Res}_{\zeta=-k}H(\zeta)}{\operatorname*{Res}_{\zeta=k}H(\zeta)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

To describe the transmitted wave, we define the function

$$G(\theta)=\frac{2\pi|f(\theta)|^2}{2b(|A|^2-|B|^2)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where $|f(\theta)|^2$ measures the energy per unit angle radiated in a direction θ, and $2b(|A|^2-|B|^2)$, the total average energy transmitted from the mouth of the channel. The function $G(\theta)$ is therefore the ratio of the time average energy radiated in a direction θ compared with the average energy which would be radiated by a simple source at the origin of the same total strength.

Equations (24), (25) and (26) also show that

$$G(\theta)=\frac{2bk|f(\theta)|^2}{|f(0)|^2(1-|R|^2)} \quad \ldots\ldots (28)$$

$$=\frac{1}{2bk}\frac{\{k\sin\theta\sin(kb\sin\theta)\}^2|H(k\cos\theta)|^2}{|\underset{\zeta=k}{\text{Res}}\,H(\zeta)|^2-|\underset{\zeta=-k}{\text{Res}}\,H(\zeta)|^2}, \quad \ldots (29)$$

which depends only on the function $H(\zeta)$.

Finally, in terms of $G(\theta)$, the amplitude of the absorbed wave, given by (14), satisfies the relation

$$|C|^2=G(\theta)/G(0). \quad \ldots\ldots (30)$$

The object of the rest of the paper is to obtain explicit expressions for the reflection coefficient R, and the function $G(\theta)$, which reduces the analysis, effectively, to the derivation of $H(\zeta)$, the transform of $H(x)$, since the relevant quantities depend on this function.

We show first that $H(x)$ satisfies an integral equation similar to the Wiener-Hopf type (Titchmarsh 1937). The usual method of solution of this equation is well adapted to the present problem since it is formulated in terms of the transform of $H(x)$, with which we are directly concerned.

The Integral Equation for $H(x)$.

The boundary conditions at our disposal in deriving $H(x)$ are those of zero normal velocity at the walls of the channel, together with the known asymptotic forms of the potential at infinity. Such problems can be solved by expressing the potential in terms of an appropriate Green's Function. It is known from Green's theorem that if ϕ is a solution of the equation $(\nabla^2+k^2)\phi=0$, continuous along with its derivatives inside a closed curve C, and if Γ is another solution, continuous except near the point P where $\Gamma\sim-\log\rho$, ρ being the distance from P, then

$$\phi(P)=\frac{1}{2\pi}\int_C\left(\Gamma\frac{\partial\phi}{\partial n}-\phi\frac{\partial\Gamma}{\partial n}\right)ds, \quad \ldots\ldots (31)$$

n being the outward drawn normal to C.

If, in vector notation, we denote by $\boldsymbol{r}$, $\boldsymbol{r}'$, the position of a general point in space, and the particular point P, respectively, then such a function satisfying the above conditions is

$$\Gamma(k|\boldsymbol{r}-\boldsymbol{r}'|)=\Gamma(x-x',\ y-y')=\frac{\pi i}{2}H_0^{(1)}(k|\boldsymbol{r}-\boldsymbol{r}'|), \quad \ldots (32)$$

where $H_0^{(1)}(z)$ is the Bessel function of the third kind as defined by Nielson (Watson 1944). We note that the asymptotic form of $H_0^{(1)}(z)$ is, for large z,

$$H_0^{(1)}(z)\sim\left(\frac{2}{\pi z}\right)^{\frac{1}{2}}\exp[i(z-\pi/4)]. \quad \ldots\ldots (33)$$

We now take the contour of integration in (31) to be the curve s_1, s_2, s_3 shown in fig. 2. By virtue of the known asymptotic behaviour of ϕ_a

and Γ, the contributions from s_1 and s_3 vanish in the limit, and since $\partial\phi_a/\partial n=0$ along the walls of the channel, we obtain, finally, the relation

$$\phi_a(r)=\int_{-\infty}^{0} \mathrm{H}(x')\frac{\partial}{\partial y'}\left[\Gamma(|\boldsymbol{r}-\boldsymbol{r}'|)_{y'=b}-\Gamma(|\boldsymbol{r}-\boldsymbol{r}'|)_{y'=-b}\right]dx', \quad . \quad (34)$$

and the boundary condition that the normal derivative of ϕ_a should vanish at the walls of the channel gives

$$\int_{-\infty}^{0} \mathrm{H}(x')\frac{\partial^2}{\partial y\partial y'}\left[\Gamma(|\boldsymbol{r}-\boldsymbol{r}'|)_{y'=b}-\Gamma(|\boldsymbol{r}-\boldsymbol{r}'|)_{y'=-b}\right]_{y=b}dx'=0, \quad . \quad (35)$$

if $x<0$.

Equation (35) is the integral equation for $\mathrm{H}(x)$. For convenience it is written

$$\int_{-\infty}^{0} \mathrm{H}(x')\mathrm{K}(x-x')\,dx'=0, \quad x<0, \quad . \quad . \quad . \quad . \quad (36)$$

where

$$\mathrm{K}(x)=-\frac{1}{2\pi}\frac{\partial^2}{\partial y\partial y'}\left[\Gamma(x, y-y')_{y'=b}-\Gamma(x, y-y')_{y'=-b}\right]_{y=b} \quad . \quad . \quad (37)$$

In deriving the above equation, no restriction to dominant mode propagation was made, and the result holds for an incident wave of arbitrary frequency. It is in determining the precise nature of the solution that the asymptotic behaviour of $\mathrm{H}(x)$ is required. In the present case the solution is subject to the condition that, for large x,

$$\mathrm{H}(x)\sim\mathrm{A}\exp[ikx]+\mathrm{B}\exp[-ikx]. \quad . \quad . \quad . \quad . \quad (38)$$

which is sufficient to determine the function uniquely.

Analogous results were obtained by Levine and Schwinger (1948) for the propagation of sound waves in a circular pipe. It was shown by these authors that integral equations of the above type could be solved by the use of Fourier transforms and their analysis is here adapted to the two-dimensional channel.

Solution of the Integral Equation.

In solving (36), the equation is first written in the more general form

$$\left.\begin{aligned}\int_{-\infty}^{0} \mathrm{H}(x')\mathrm{K}(x-x')\,dx'&=\mathrm{E}(x), \quad x\geqslant 0\\ &=0, \quad x<0,\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (39)$$

so that, multiplying each side by $\exp[-i\zeta x]$, and integrating over all x, we obtain

$$\int_{-\infty}^{0} \mathrm{H}(x')\,dx'\int_{-\infty}^{\infty} \mathrm{K}(x-x')\exp[-i\zeta x]\,dx=\int_{0}^{\infty} \mathrm{E}(x)\exp[-i\zeta x]\,dx,$$

or

$$\int_{-\infty}^{0} \mathrm{H}(x')\exp[-i\zeta x']\,dx'\int_{-\infty}^{\infty} \mathrm{K}(z)\exp[-i\zeta z]\,dz=\int_{0}^{\infty} \mathrm{E}(x)\exp[-i\zeta x]\,dx,$$

i. e.

$$\mathrm{H}(\zeta)\mathrm{K}(\zeta)=\mathrm{E}(\zeta), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

$\mathrm{E}(\zeta)$ being the transform of the right-hand side of (39).

In the application of equation (40) it is necessary that the functions occurring should have a common region of regularity. We show that this is true for $|\eta|<\epsilon$ where $\zeta=\xi+i\eta$ and $Imk=\epsilon$.

From the asymptotic form of $\mathrm{H}(x)$ we know that $\mathrm{H}(\zeta)$ is regular for $\eta>\epsilon$ but has simple poles at $\zeta=\pm k$. However, by inserting the factor (ζ^2-k^2), we may take $(\zeta^2-k^2)\mathrm{H}(\zeta)$ to be regular for $\eta\geqslant-\epsilon$.

Moreover, by inserting the asymptotic form of Γ in equation (37) we obtain

$$\mathrm{K}(x)\sim\left(\frac{\pi}{2k}\right)^{\frac{1}{2}}\frac{\exp\left[i(k|x|+\pi)\right]}{|x|^{\frac{3}{2}}} \quad . \quad . \quad . \quad . \quad . \quad (41)$$

so that $\mathrm{K}(\zeta)$ is regular for $|\eta|<\epsilon$.

The explicit relation for $\mathrm{K}(\zeta)$ can be obtained from that of $\Gamma(\zeta, y-y')$, which is shown in Appendix I. to be

$$\Gamma(\zeta, y-y')=\frac{i\pi\exp\left[i(k^2-\zeta^2)^{\frac{1}{2}}|y-y'|\right]}{(k^2-\zeta^2)^{\frac{1}{2}}}, \quad . \quad . \quad . \quad . \quad (42)$$

where the argument of $(k^2-\zeta^2)^{\frac{1}{2}}$ is taken to be $\pi/2$ as $|\zeta|\to\infty$ on the real axis.

It follows, from equation (37), that

$$\mathrm{K}(\zeta)=b(k^2-\zeta^2)\mathrm{L}(\zeta), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

where

$$\mathrm{L}(\zeta)=-\frac{\left[1-\exp\left\{2bi(k^2-\zeta^2)^{\frac{1}{2}}\right\}\right]}{2bi(k^2-\zeta^2)^{\frac{1}{2}}} \quad . \quad . \quad . \quad . \quad (44)$$

and $\mathrm{L}(\zeta)$ is regular for $|\eta|<\epsilon$. Moreover $\mathrm{L}(\zeta)$ is non-zero inside the strip ; its zeros occur at $\zeta=\pm i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}$ and these points lie outside the strip.

Thus the left-hand side of (40) is composed of two factors, $\mathrm{L}(\zeta)$ and $b(k^2-\zeta^2)\mathrm{H}(\zeta)$, for which $|\eta|<\epsilon$ is a common region of regularity.

Finally we note, from (34) and (37) and the definition of $\mathrm{E}(x)$ contained in (39), that $\mathrm{E}(x)$ represents the y-derivative of the potential at $y=b$. Its asymptotic behaviour can therefore be derived from that of ϕ_a outside the channel. With the help of (4) we obtain, for large x,

$$\mathrm{E}(x)\sim-\tfrac{1}{2}bf(0)\frac{\exp\left[ikx\right]}{x^{3/2}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

so that $\mathrm{E}(\zeta)$ is regular for $\eta\leqslant\epsilon$.

The next step in the solution of (39) is to identify $\mathrm{L}(\zeta)$ as the quotient of two functions, $\mathrm{L}_+(\zeta)$ and $\mathrm{L}_-(\zeta)$, which are regular and non-zero for $\eta>-\epsilon$, $\eta<\epsilon$, respectively. Equation (40) will then become

$$b(k^2-\zeta^2)\mathrm{H}(\zeta)\mathrm{L}_+(\zeta)=\mathrm{E}(\zeta)\mathrm{L}_-(\zeta), \quad . \quad . \quad . \quad . \quad . \quad (46)$$

the left-hand side being regular for $\eta>-\epsilon$ and the right-hand side for

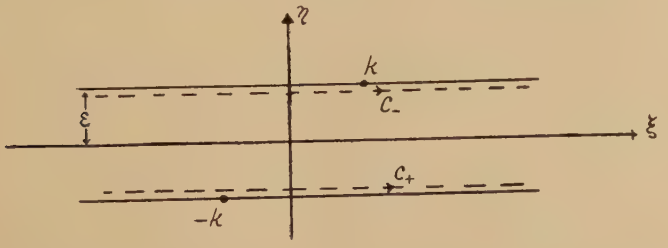
$\eta<\epsilon$. By analytic continuation they define, together, a function regular in the whole finite domain. It will be shown in Appendix II. that the integral function so defined is a constant so that a knowledge of $L_+(\zeta)$ implies a knowledge of $H(\zeta)$ apart from a constant factor. Reference to equations (26) and (29) shows that the expressions for the reflection coefficient R, and the function $G(\theta)$ describing the transmitted energy, are independent of such a factor multiplying $H(\zeta)$ so that its explicit evaluation is unnecessary.

The functions $L_+(\zeta)$ and $L_-(\zeta)$ are obtained by applying Cauchy's formula to $\log L(\zeta)$. Since $L(\zeta)$ is regular and non-zero in the region $|\eta|<\epsilon$, we can write, for $|\eta|<\epsilon$,

$$\log L(\zeta)=\frac{1}{2\pi i}\oint\frac{\log L(t)\,dt}{t-\zeta}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

the integration being taken round an elongated rectangle within the strip and on the branch for which $\log 1=0$ (fig. 3).

Fig. 3.



From (44) we find that $\log L(t)\sim\log|t|$ as $|t|\to\infty$ within the strip. Consequently the contributions from the vertical sides of the rectangle tend to zero as the contour extends to infinity and we have the relation

$$\log L(\zeta)=\frac{1}{2\pi i}\int_{C_+}\frac{\log L(t)\,dt}{t-\zeta}-\frac{1}{2\pi i}\int_{C_-}\frac{\log L(t)}{t-\zeta} \quad . \quad . \quad (46)$$

$$=\log L_+(\zeta)-\log L_-(\zeta). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

The contours C_+ and C_- are shown in fig. 3, and, to ensure the individual convergence of L_+ and L_- we take

$$\int_{C_+}\frac{\log L(t)\,dt}{t-\zeta}=\lim_{T\to\infty}\int_{-T-i\delta}^{T-i\delta}\frac{\log L(t)\,dt}{t-\zeta}, \quad . \quad . \quad . \quad . \quad (48)$$

where $0<\delta<\epsilon$.

The second integral is defined in a similar manner.

The two functions $L_+(\zeta)$ and $L_-(\zeta)$ so defined are regular and non-zero in the regions $\eta>-\delta$ and $\eta<\delta$ respectively. In the limit when the two

contours approach the axis, the two functions L_+ and L_- are represented by the common expression

$$\exp\left[\frac{1}{2\pi i}\lim_{T\to\infty}\int_{-T}^{T}\frac{\log L(t)\,dt}{t-\zeta}\right]$$

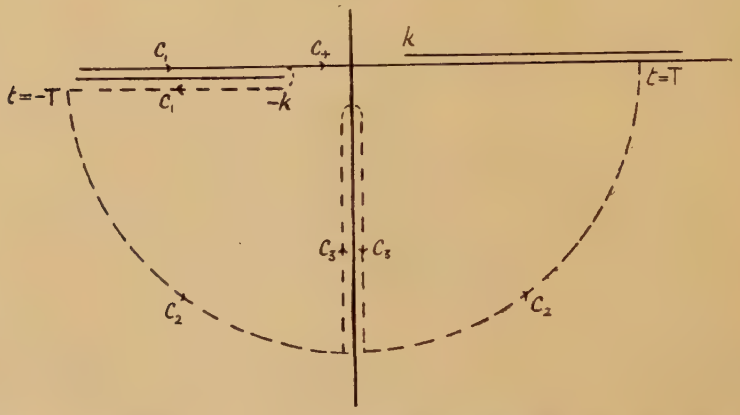
$$=\exp\left[\frac{1}{2\pi i}\lim_{T\to\infty}\int_{-T}^{T}\log\left\{\frac{i\{1-\exp[2bi(k^2-t^2)^{\frac{1}{2}}]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{t-\zeta}\right] \quad . \quad . \quad (49)$$

and

$$L_+(-\zeta)=\frac{1}{L_-(\zeta)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

For real k (*i. e.* as $\epsilon\to 0$), the contours C_+, C_- should pass the branch point $t=k$ on the lower side of the real axis and $t=-k$ on the upper side. The two contours then lie wholly in the respective regions of regularity of the two functions L_+, L_- and the phase of $(k^2-t^2)^{\frac{1}{2}}$ changes from 0 to $\pi/2$ for $|t|<k$, $|t|>k$ respectively.

Fig. 4.



The above expression for L_+ could now be used to deduce the function $H(\zeta)$ apart from a constant factor, and hence the asymptotic behaviour of the reflected and transmitted waves explicitly. It is possible, however, to express $L_+(\zeta)$ in a somewhat more significant form which is most easily obtained by a deformation of the contour C_+. The discussion of convergence is also simplified if we proceed from the derivative of $\log L_+(\zeta)$ which is

$$\frac{d}{d\zeta}\log L_+(\zeta)=\frac{1}{2\pi i}\int_{C_+}\log\left\{\frac{i\{1-\exp[2bi(k^2-t^2)^{\frac{1}{2}}]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{(t-\zeta)^2}. \quad . \quad . \quad (51)$$

Remembering that ζ lies in the upper half plane, we deform the contour C_+ into the contour $C_1\,C_2\,C_3$ shown in fig. 4. The section C_1 is the upper and lower sides of the branch line extending from $-\infty$ to $-k$; C_2 consists of two arcs of a circle, radius T, in the lower half-plane, while C_3 excludes the singularities of the integrand in (51) at the zeros of $\{1-\exp[2bi(k^2-t^2)^{\frac{1}{2}}]\}$ on the imaginary axis.

Along the upper and lower sides of the branch cut, the argument of $(k^2-t^2)^{\frac{1}{2}}$ is $+\pi/2$ and $-\pi/2$ respectively, so that the contribution to the right-hand side of equation (51) from C_1 is

$$\frac{1}{2\pi i}\int_{C_1}\log\left\{\frac{i\{1-\exp[2bi(k^2-t^2)^{\frac{1}{2}}]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{(t-\zeta)^2}$$

$$=\frac{1}{2\pi i}\int_k^{\mathrm{T}}\log\left[\frac{1-\exp[-2b(t^2-k^2)^{\frac{1}{2}}]}{-(1-\exp[-2b(t^2-k^2)^{\frac{1}{2}}])}\right]\frac{dt}{(t+\zeta)^2}=-\frac{b}{\pi i}\int_k^{\mathrm{T}}\frac{(t^2-k^2)^{\frac{1}{2}}\,dt}{(t+\zeta)^2}$$

$$=-\frac{b}{\pi i}\left[-\frac{(\mathrm{T}^2-k^2)^{\frac{1}{2}}}{\mathrm{T}+\zeta}+\cosh^{-1}\mathrm{T}/k-\zeta\int_0^{\cosh^{-1}\mathrm{T}/k}\frac{du}{k\cosh u+\zeta}\right] \quad . \quad . \quad . \quad (52)$$

so that

$$\lim_{\mathrm{T}\to\infty}\left[\frac{1}{2\pi i}\int_{C_1}+\frac{b}{\pi i}\cosh^{-1}\frac{\mathrm{T}}{k}\right]=\frac{b}{\pi i}\left[1+\frac{2\zeta}{(k^2-\zeta^2)^{\frac{1}{2}}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}\right] \quad . \quad . \quad (53)$$

Consider now the integral round C_2 in the quadrant for which $-\pi<\arg t<-\pi/2$. In this range $-\pi/2<\arg(k^2-t^2)^{\frac{1}{2}}<0$, and the asymptotic value of the integrand in (51) is $-(\log 2bt)/t^2$ so that, in the limit $\mathrm{T}\to\infty$, the integral along C_2 is zero in this quadrant.

In the range $-\pi/2<t<0$, however, we have $0<\arg(k^2-t^2)^{\frac{1}{2}}<\pi/2$. On this section of C_2 the asymptotic value of the integrand is $(\log\exp[-2bt])/t^2=-2b/t$ and the contribution is therefore

$$-\frac{2b}{2\pi i}\int\frac{dt}{t}=-\frac{b}{\pi i}\int_{-\pi/2}^{0}d\theta=-\frac{b}{2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

Finally, there is the contribution from C_3 due to the phase difference of the integrand on the two sides of the imaginary axis. Since the zeros of $[1-\exp\{2bi(k^2-t^2)^{\frac{1}{2}}\}]$ occur at $t=-i(n^2\pi^2/b^2-k^2)^{\frac{1}{2}}$, it follows that the integrand in (51) has the same phase, along that part of the negative real axis for which $|Imt|<\mathrm{T}$, as

$$\prod_{n=1}^{\mathrm{N}}\left[t+i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}\right],$$

where N is the largest integer not exceeding $\mathrm{T}b/\pi$. It follows that the integral along the two sides of the negative real axis is

$$\frac{1}{2\pi i}\int_{C_3}\log\left\{\frac{i\{1-\exp[2bi(k^2-t^2)^{\frac{1}{2}}]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{(t-\zeta)^2}$$

$$=\frac{1}{2\pi i}\int_{C_3}\log\prod_{n=1}^{\mathrm{N}}\left\{t+i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}\right\}\frac{dt}{(t-\zeta)^2}$$

$$=-\sum_{n=1}^{\mathrm{N}}\int_{-i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}^{-i\mathrm{T}}\frac{dt}{(t-\zeta)^2}=-\sum_{n=1}^{\mathrm{N}}\left[\frac{1}{\zeta+i\mathrm{T}}-\frac{1}{\zeta+i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}\right]$$

$$. \quad . \quad . \quad (55)$$

If the contributions from the various sections of contour, given by (53), (54), (55), are now added, we obtain

$$\frac{d}{d\zeta}\log \mathrm{L}_+(\zeta)=\frac{b}{\pi i}\left[1+\frac{2\zeta}{(k^2-\zeta^2)^{\frac{1}{2}}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}\right]-\frac{b}{2}$$
$$-\lim_{\mathrm{T}\to\infty}\left[\sum_{n=1}^{\mathrm{N}}\left\{\frac{1}{\zeta+i\mathrm{T}}-\frac{1}{\zeta+i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}+\frac{b}{\pi i}\cosh^{-1}\frac{\mathrm{T}}{k}\right\}\right]. \quad . \quad . \quad . \quad (56)$$

Now, since, for large z and N,

$$\cosh^{-1}z\sim\log 2z$$

$$\sum_{n=1}^{\mathrm{N}}\frac{1}{n}\sim\log \mathrm{N}+\gamma,$$

where γ is Euler's constant, it follows that (56) can be written

$$\frac{d}{d\zeta}\log \mathrm{L}_+(\zeta)=\frac{b}{\pi i}\left[1+\frac{2\zeta}{(k^2-\zeta^2)^{\frac{1}{2}}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}\right]-\frac{b}{2}$$
$$+\lim_{\mathrm{T}\to\infty}\left[\sum_{n=1}^{\mathrm{N}}\left\{\frac{1}{\zeta+i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{1}{\zeta+i\mathrm{T}}-\frac{b}{n\pi i}\right\}\right.$$
$$\left.+\frac{b\gamma}{\pi i}+\frac{b}{\pi i}\log \mathrm{N}-\frac{b}{\pi i}\log\frac{2\mathrm{T}}{k}\right]=\frac{2\zeta b}{\pi i(k^2-\zeta^2)^{\frac{1}{2}}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}-\frac{b}{2}$$
$$+\sum_{n=1}^{\infty}\left\{\frac{1}{\zeta+i\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{b}{n\pi i}\right\}+\frac{b\gamma}{\pi i}+\frac{b}{\pi i}\log\frac{bk}{2\pi}, \quad . \quad (57)$$

since $\mathrm{N}\sim \mathrm{T}b/\pi$.

A simple integration with respect to ζ then gives

$$\mathrm{L}_+(\zeta)=\kappa\exp\left[-\frac{b\zeta}{2}+\frac{i\zeta b}{\pi}\log\frac{2\pi}{bk}-\frac{i\zeta b\gamma}{\pi}+\frac{i\zeta b}{\pi}+\frac{2ib}{\pi}(k^2-\zeta^2)^{\frac{1}{2}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}\right.$$
$$\left.+\sum_{n=1}^{\infty}\left\{\log\left[\frac{\zeta b}{n\pi}+i\left(1-\frac{k^2b^2}{n^2\pi^2}\right)^{\frac{1}{2}}\right]+\frac{ib\zeta}{n\pi}\right\}\right]$$
$$=\kappa'\prod_{n=1}^{\infty}\left[\left(1-\frac{k^2b^2}{n^2\pi^2}\right)^{\frac{1}{2}}-\frac{i\zeta b}{n\pi}\right]\exp\left(\frac{i\zeta b}{n\pi}\right)\exp\left[\frac{i\zeta b}{\pi}\left(1-\gamma+\log\frac{2\pi}{bk}\right)-\frac{b\zeta}{2}\right.$$
$$\left.+\frac{2ib}{\pi}(k^2-\zeta^2)^{\frac{1}{2}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}\right]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (58)$$

Since

$$\mathrm{L}_+(-\zeta)=\frac{1}{\mathrm{L}_-(\zeta)},$$

equation (58) gives

$$\mathrm{L}_+(\zeta)/\mathrm{L}_-(\zeta)=\mathrm{L}(\zeta)=(\kappa')^2\prod_{n=1}^{\infty}\left[1-\frac{(k^2-\zeta^2)b^2}{n^2\pi^2}\right]\exp\left[ib(k^2-\zeta^2)^{\frac{1}{2}}\right]$$
$$=\frac{(\kappa')^2 i}{2b(k^2-\zeta^2)^{\frac{1}{2}}}\left(1-\exp\left[2bi(k^2-\zeta^2)^{\frac{1}{2}}\right]\right), \quad . \quad . \quad . \quad (59)$$

and comparison of equation (59) with (44) shows that the integration constant $\kappa'=1$.

Moreover

$$\prod_{n=1}^{\infty}\left\{\left(1-\frac{k^2b^2}{n^2\pi^2}\right)^{\frac{1}{2}}-\frac{i\zeta b}{n\pi}\right\}\exp\left[\frac{i\zeta b}{n\pi}\right]$$

$$=\exp\left[\sum_{n=1}^{\infty}\log\left\{\left(1-\frac{k^2b^2}{n^2\pi^2}\right)^{\frac{1}{2}}-\frac{i\zeta b}{n\pi}\right\}+\frac{i\zeta b}{n\pi}\right]$$

$$=\exp\left[\tfrac{1}{2}\sum_{n=1}^{\infty}\log\left\{1-\frac{(k^2-\zeta^2)b^2}{n^2\pi^2}\right\}-i\sum_{n=1}^{\infty}\left\{\tan^{-1}\frac{\zeta}{\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{\zeta b}{n\pi}\right\}\right]$$

$$=\prod_{n=1}^{\infty}\left\{1-\frac{(k^2-\zeta^2)b^2}{n^2\pi^2}\right\}^{\frac{1}{2}}\exp\left[-i\sum_{n=1}^{\infty}\left\{\tan^{-1}\frac{\zeta}{\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{\zeta b}{n\pi}\right\}\right]$$

$$=\left[\frac{\sin b(k^2-\zeta^2)^{\frac{1}{2}}}{b(k^2-\zeta^2)^{\frac{1}{2}}}\right]^{\frac{1}{2}}\exp\left[-i\sum_{n=1}^{\infty}\left\{\tan^{-1}\frac{\zeta}{\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{\zeta b}{n\pi}\right\}\right], \quad . \quad . \quad . \quad (60)$$

so that

$$\mathrm{L}_+(\zeta)=\left[\frac{\sin b(k^2-\zeta^2)^{\frac{1}{2}}}{b(k^2-\zeta^2)^{\frac{1}{2}}}\right]^{\frac{1}{2}}\exp\left[-\frac{b\zeta}{2}+\frac{i\zeta b}{\pi}\left(1-\gamma+\log\frac{2\pi}{bk}\right)\right.$$

$$\left.+\frac{2ib}{\pi}(k^2-\zeta^2)^{\frac{1}{2}}\tan^{-1}\left(\frac{k-\zeta}{k+\zeta}\right)^{\frac{1}{2}}-i\sum_{n=1}^{\infty}\left\{\tan^{-1}\frac{\zeta}{\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{\zeta b}{n\pi}\right\}\right], \quad . \quad . \quad . \quad (61)$$

and

$$\mathrm{L}_+(k\cos\theta)=\left[\frac{\sin(bk\sin\theta)}{bk\sin\theta}\right]^{\frac{1}{2}}\exp\left[-\frac{bk\cos\theta}{2}+\frac{ibk\cos\theta}{\pi}\left(1-\gamma+\log\frac{2\pi}{bk}\right)\right.$$

$$\left.+\frac{ibk\theta\sin\theta}{\pi}-i\sum_{n=1}^{\infty}\left\{\tan^{-1}\frac{k\cos\theta}{\left(\frac{n^2\pi^2}{b^2}-k^2\right)^{\frac{1}{2}}}-\frac{bk\cos\theta}{n\pi}\right\}\right]. \quad . \quad (62)$$

$$\mathrm{L}_+(k)=\exp\left[-\frac{bk}{2}+\frac{ibk}{\pi}\left(1-\gamma+\log\frac{2\pi}{bk}\right)-i\sum_{n=1}^{\infty}\left\{\sin^{-1}\frac{bk}{n\pi}-\frac{bk}{n\pi}\right\}\right]. \quad . \quad (63)$$

It is now a simple matter to obtain expressions describing the reflected and transmitted waves. The reflection coefficient is, from equation (26),

$$\mathrm{R}=\operatorname*{Res}_{\zeta=-k}\mathrm{H}(\zeta)\Big/\operatorname*{Res}_{\zeta=k}\mathrm{H}(\zeta), \quad . \quad . \quad . \quad . \quad . \quad . \quad (64)$$

and since, (see Appendix)

$$(k^2-\zeta^2)\mathrm{H}(\zeta)\mathrm{L}_+(\zeta)=\mathrm{D}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (65)$$

where D is some constant, it follows that

$$\operatorname*{Res}_{\zeta=\pm k} = \pm \frac{\mathrm{D}}{2k\mathrm{L}_+(\pm k)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (66)$$

Equation (64) can now be written

$$\mathrm{R}=-|\,\mathrm{R}\,|\exp[2ikl]=-\mathrm{L}_+(k)/\mathrm{L}_+(-k)=-\{\mathrm{L}_+(k)\}^2, \quad . \quad (67)$$

where l is the end correction.

Using the expression for $\mathrm{L}_+(k)$ given in equation (63), we find that

$$|\,\mathrm{R}\,|=\exp[-bk], \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (68)$$

$$\frac{l}{b}=\frac{1}{\pi}\left(1-\gamma+\log\frac{2\pi}{bk}\right)-\frac{1}{bk}\sum_{n=1}^{\infty}\left\{\sin^{-1}\frac{bk}{n\pi}-\frac{bk}{n\pi}\right\}. \quad . \quad . \quad (69)$$

Finally we derive the expression for $\mathrm{G}(\theta)$. Equation (65) shows that

$$\mathrm{H}(k\cos\theta)=\frac{\mathrm{D}}{(k\sin\theta)^2\mathrm{L}_+(k\cos\theta)}. \quad . \quad . \quad . \quad . \quad . \quad (70)$$

Substituting (66) and (70) in (29) we obtain

$$\mathrm{G}(\theta)=\frac{2\sin^2(bk\sin\theta)}{bk\sin^2\theta}\frac{|\,\mathrm{L}_+(k)\,|^2}{(1-|\,\mathrm{R}\,|)^2|\,\mathrm{L}_+(k\cos\theta)\,|^2} \quad . \quad . \quad . \quad (71)$$

$$=\frac{\exp[bk\cos\theta]\sin(bk\sin\theta)}{\sin\theta\sinh bk}, \quad . \quad . \quad . \quad . \quad . \quad (72)$$

by virtue of relations (62), (63) and (68).

A simple check on this last relation is obtained by noting that from its definition the function $\mathrm{G}(\theta)$ must satisfy

$$\int_{-\pi}^{\pi}\mathrm{G}(\theta)\,d\theta=2\pi. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (73)$$

To verify this we consider the integral

$$2\oint\frac{\exp[bkz]}{z^2-1}dz, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (74)$$

taken round the unit circle in the z-plane, with indentations at the poles $z=\pm 1$.

The integrand is free from singularities inside the contour and the integral is therefore zero. If we substitute $z=\exp[i\theta]$, then this is equivalent to

$$\mathrm{P}\int_{-\pi}^{\pi}\exp[bk\exp(i\theta)]\frac{d\theta}{\sin\theta}-\pi i\left[\operatorname*{Res}_{z=1}+\operatorname*{Res}_{z=-1}\right]=0 \quad . \quad . \quad (75)$$

or

$$\int_{-\pi}^{\pi} \exp\,[bk\cos\theta]\,\frac{\sin\,(bk\sin\theta)}{\sin\theta}\,d\theta=\pi(\exp\,[bk]+\exp\,[-bk]) \qquad (76)$$

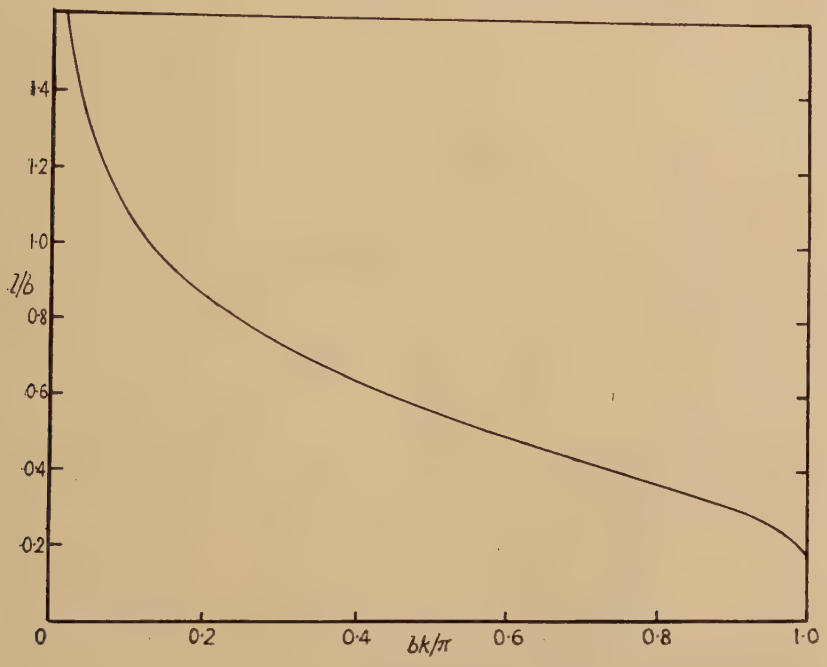
so that

$$\int_{-\pi}^{\pi} \mathrm{G}(\theta)\,d\theta=\int_{-\pi}^{\pi} \exp\,[bk\cos\theta]\,\frac{\sin\,(bk\sin\theta)}{\sin\theta\,\sinh bk}\,d\theta=2\pi. \quad . \quad . \quad (77)$$

Discussion of Results.

One of the most striking features is the extremely simple relation for the amplitude of the reflected wave (equation (68)). The corresponding

Fig. 5.



The variation in end-correction as a function of $\frac{bk}{\pi}\left(=\frac{2b}{\lambda}\right)$.

result for the circular pipe is more complex, but, for large values of the wavelength, the amplitude is approximately proportional to $\exp\,[-k^2c^2/2]$ where c is the radius of the pipe (Levine and Schwinger 1948). The decrease amplitude with the wavelength, when the latter is large, is therefore smaller than in the two-dimensional case.

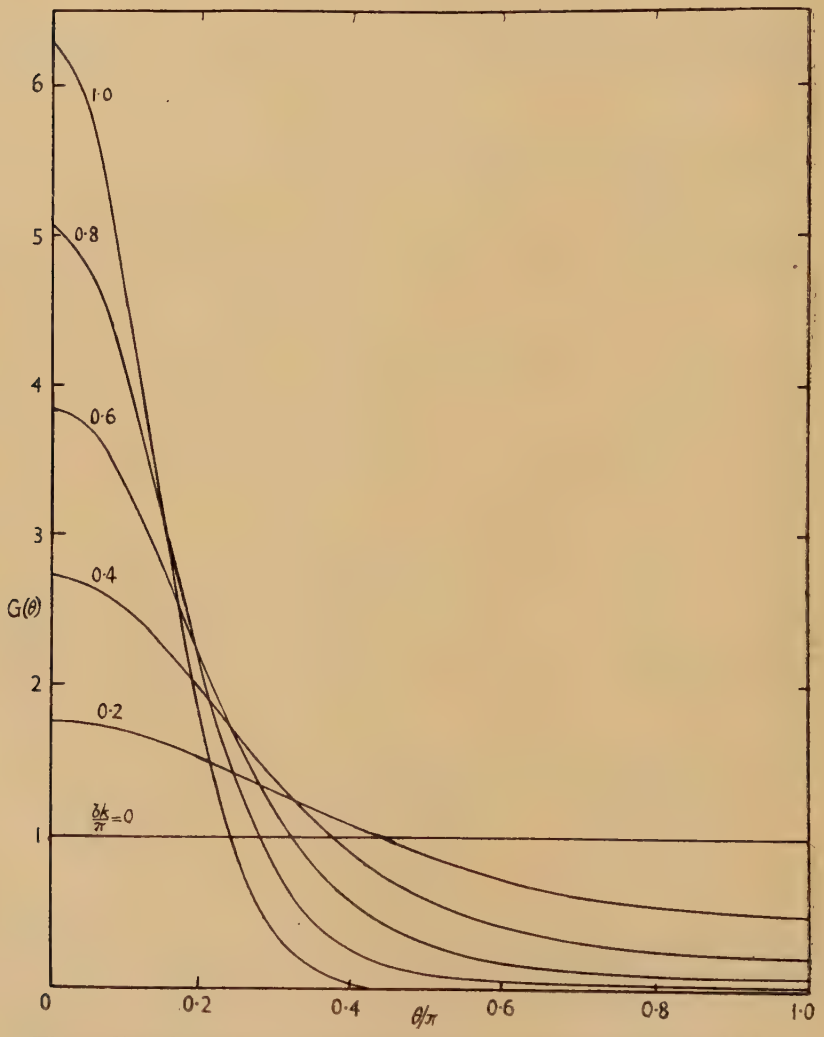
The end correction is illustrated graphically in fig. 5. Equation (69) shows that it increases indefinitely as the wavelength becomes infinitely great, as opposed to the circular pipe for which it tends to a finite limit. The same conclusion was reached by Rayleigh (1904) using approximate

methods. The physical significance of the end-correction is consequently somewhat remote and it is more constructive to consider the actual correction to the phase in the returning wave which is given by

$$2kl = \frac{2bk}{\pi}\left(1-\gamma+\log\frac{2\pi}{bk}\right)-2\sum_{n=1}^{\infty}\left\{\sin^{-1}\frac{bk}{n\pi}-\frac{bk}{n\pi}\right\} \quad . \quad . \quad (78)$$

and tends to zero for large values of the wavelength.

Fig. 6.



The radiated energy as a function of θ/π for various values of bk/π.

In fig. 6, the variation in $G(\theta)$ is shown for various values of $bk/\pi(=2b/\lambda)$. When the wavelength is comparable with the width of the channel, most of the energy is transmitted in a forward direction. The function

decreases monotonically with θ except for values of bk/π near unity when there is a minimum in the region of $\theta=\pi/2$. If $bk=\pi$, for example, $G(\pi/2)=0$ and no energy is propagated in this direction. Beyond this value $G(\theta)$ increases slowly to 0·0118.

Appendix I.

To derive the Fourier transform of

$$\Gamma(x, y)=\frac{\pi i}{2}\,H_0^{(1)}[k(x^2+y^2)^{\frac{1}{2}}] \quad . \quad . \quad . \quad . \quad . \quad (A.\ 1)$$

we begin from the known relation (Watson 1944)

$$H_0^{(1)}(z)=\frac{1}{\pi i}\int_{-\infty}^{\infty}\exp\,[iz\cosh t]\,dt \text{ if } 0\leqslant\arg z\leqslant\pi, \quad . \quad . \quad (A.\ 2)$$

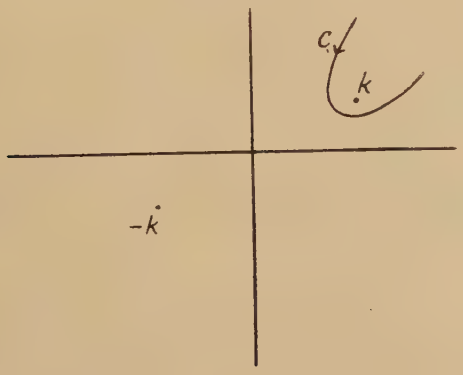
so that

$$\Gamma(x, y)=\tfrac{1}{2}\int_{-\infty}^{\infty}\exp\,[ik(x^2+y^2)^{\frac{1}{2}}\cosh t]\,dt \text{ if } Imk=\epsilon>0$$

$$=\tfrac{1}{2}\int_{-\infty-i\theta}^{\infty-i\theta}\exp\,[ikx\cosh u-k|y|\sinh u]\,du, \quad . \quad (A.\ 3)$$

where $\tan\theta=|y|/z$.

Fig. 7.



Substituting $\zeta=k\cosh u$ this becomes

$$\Gamma(x, y)=-\frac{1}{2i}\int\frac{\exp[ix\zeta+i|y|(k^2-\zeta^2)^{\frac{1}{2}}]\,d\zeta}{(k^2-\zeta^2)^{\frac{1}{2}}}, \quad . \quad . \quad (A.\ 4)$$

where the contour of integration begins and ends at infinity and passes between the two branch points $\zeta=\pm k$ (see fig. 7). By a deformation of the contour, this can be written, if $|\eta|<\epsilon$

$$\Gamma(x, y)=-\frac{1}{2i}\int_{-\infty+i\eta}^{\infty+i\eta}\frac{\exp\,[ix\zeta+i|y|(k^2-\zeta^2)^{\frac{1}{2}}]\,d\zeta}{(k^2-\zeta^2)^{\frac{1}{2}}}. \quad . \quad (A.\ 5)$$

It follows that the transform of $\Gamma(x, y)$ is $\Gamma(\zeta, y)$ where

$$\Gamma(\zeta, y)=\frac{i\pi \exp [i(k^2-\zeta^2)^{\frac{1}{2}} \mid y \mid]}{(k^2-\zeta^2)^{\frac{1}{2}}} \quad . \quad . \quad . \quad . \quad . \quad (\text{A. } 6)$$

and is regular for $|\mathrm{I}m\zeta|<\epsilon$.

Appendix II.

During the course of the analysis, use was made of the fact that the integral function defined by the two sides of the equation

$$b(k^2-\zeta^2)^{\frac{1}{2}}\mathrm{H}(\zeta)\mathrm{L}_{+}(\zeta)=\mathrm{E}(\zeta)\mathrm{L}_{-}(\zeta)$$

was, in fact, a constant.

To prove this we examine the asymptotic values of the two sides, in their respective regions of regularity, as $|\zeta| \rightarrow \infty$.

For the asymptotic behaviour of $\mathrm{L}_{+}(\zeta)$ we refer to equation (51),

$$\begin{aligned}\mathrm{L}_{+}(\zeta)&=\exp\left[\frac{1}{2\pi i}\lim_{\mathrm{T}\rightarrow\infty}\int_{-\mathrm{T}}^{\mathrm{T}}\log\left\{\frac{i\{1-\exp [2bi(k^2-t^2)^{\frac{1}{2}}]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{t-\zeta}\right]\\ &=\exp\left[\frac{1}{2\pi i}\int_{-k}^{k}\log\left\{\frac{i\{1-\exp [2bi(k^2-t^2)^{\frac{1}{2}}]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{t-\zeta}\right.\\ &\quad\left.+\frac{\zeta}{\pi i}\int_{k}^{\infty}\log\left\{\frac{1-\exp [-2b(t^2-k^2)^{\frac{1}{2}}]}{2b(t^2-k^2)^{\frac{1}{2}}}\right\}\frac{dt}{t^2-\zeta^2}\right]. \quad . \quad . \quad . \quad (\text{A. } 7)\end{aligned}$$

Now

$$\frac{1}{2\pi i}\int_{-k}^{k}\log\left\{\frac{i\{1-\exp [2bi(k^2-t^{2\frac{1}{2}})]\}}{2b(k^2-t^2)^{\frac{1}{2}}}\right\}\frac{dt}{t-\zeta}=-0\left(\frac{1}{\zeta}\right) \quad . \quad (\text{A. } 8)$$

and

$$\begin{aligned}\int_{k}^{\infty}\log\left\{\frac{1-\exp [-2b(t^2-k^2)^{\frac{1}{2}}]}{2b(t^2-k^2)^{\frac{1}{2}}}\right\}\frac{dt}{t^2-\zeta^2}&=\int_{k}^{\infty}\log\left\{\frac{t\{1-\exp [-2b(t^2-k^2)^{\frac{1}{2}}]\}}{2b(t^2-k^2)^{\frac{1}{2}}}\right\}\\ &\quad\times\frac{dt}{t^2-\zeta^2}+\int_{k}^{\infty}\log\left(\frac{1}{t}\right)\frac{dt}{t^2-\zeta^2}\\ &=-0\left(\frac{1}{\zeta^2}\right)+\frac{i}{\zeta}\int_{-k/i\zeta}^{\infty}\log\left(\frac{1}{-i\zeta v}\right)\frac{dv}{1+v^2}\\ &\qquad . \quad . \quad . \quad (\text{A. } 9)\\ &\sim\frac{\pi i}{2\zeta}\log\left(\frac{1}{-i\zeta}\right)+0\left(\frac{1}{\zeta}\right). \quad . \quad (\text{A. } 10)\end{aligned}$$

Substituting in (A. 7) we obtain

$$\mathrm{L}_{+}(\zeta)\sim\frac{1}{(-i\zeta)^{\frac{1}{2}}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{A. } 11)$$

and

$$\mathrm{L}_{-}(\zeta)=\frac{1}{\mathrm{L}_{+}(-\zeta)}\sim(i\zeta)^{\frac{1}{2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{A. } 12)$$

For the asymptotic behaviour of

$$\mathrm{H}(\zeta)=\int_{-\infty}^{0}\mathrm{H}(x)\exp\left[-i\zeta x\right]dx \quad Im\zeta>0, \quad . \quad . \quad \text{(A. 13)}$$

and

$$\mathrm{E}(\zeta)=\int_{0}^{\infty}\mathrm{E}(x)\exp\left[-i\zeta x\right]dx \quad Im\zeta<0, \quad . \quad . \quad . \quad \text{(A. 14)}$$

we note that the significant contributions to the integrals, when $|\zeta|$ is large, will arise from the values of $\mathrm{H}(x)$, $\mathrm{E}(x)$ for small negative and positive x respectively. Now the difference in potential across the plane $y=b$ must be a continuous function of x which vanishes for $x>0$. It follows that, as $x\to-0$

$$\mathrm{H}(x)\sim x^{\alpha}, \quad \alpha>0 \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(A. 15)}$$

or

$$\mathrm{H}(\zeta)\sim(-i\zeta)^{-\alpha-1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(A. 16)}$$

Again the y-derivative of the potential across the plane $y=b$ must have an integrable singularity at $x=0$, and since this is just $\mathrm{E}(x)$, $x>0$, we have, as $x\to+0$

$$\mathrm{E}(x)\sim x^{-\beta}, \quad \beta<1, \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(A. 17)}$$

or

$$\mathrm{E}(\zeta)\sim(i\zeta)^{\beta-1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(A. 18)}$$

Using (A. 11), (A. 12), (A. 13) and (A. 14) we see that the behaviour of the integral function is, in the upper half plane

$$b(k^2-\zeta^2)\mathrm{H}(\zeta)\mathrm{L}_{+}(\zeta)\sim(-i\zeta)^{\frac{1}{2}-\alpha},$$

while in the lower half plane

$$\mathrm{E}(\zeta)\mathrm{L}_{-}(\zeta)\sim(i\zeta)^{\beta-\frac{1}{2}}.$$

Because of the limits on α and β, the integral function cannot tend to infinity faster than $\zeta^{\frac{1}{2}}$ and must therefore be a constant.

References.

Carslaw, 1898, *Proc. Lond. Math. Soc.*, **30**, 121.
Levine and Schwinger, 1948, *Phys. Rev.*, **73**, 383.
Rayleigh, 1904, *Phil. Mag.*, **8**, 481.
Rayleigh, 1940, *Theory of Sound* (London : Macmillan & Co.), **2**, chap. XVI.
Sommerfeld, 1894, *Math. Ann.*, **47**, 368.
Titchmarsh, 1937, *Theory of Fourier Integrals* (Oxford : University Press).
Watson, 1944, *Theory of Bessel Functions* (Cambridge : University Press).

III. *Betatron-Starting in Electron Synchrotrons.*

By J. J. Wilkins,
Atomic Energy Research Establishment, Harwell*.

[Received September 19, 1949.]

Abstract.

The magnetic design of betatron-starting flux bars for electron synchrotrons is considered in relation to the principles governing efficient initial electron injection and efficient betatron-synchrotron transition. A brief discussion of betatron-synchrotron transition theory is included, in order to show how flux bar performance may be specified therefrom. A design procedure is derived for determining the essential flux bar dimensions, approximate corrections being made for skin effect in the laminated material of the bars.

To illustrate the principles discussed in the main paper, the results of some detailed performance calculations on the 375 MeV. Glasgow University synchrotron are given in an Appendix.

§1. Introduction.

In designing the magnet of a betatron-started electron synchrotron, economy in weight and space dictates the use of betatron flux bars of the minimum size consistent with satisfactory performance. Essentially this implies that the betatron acceleration phase of the machine is restricted to imparting just sufficient electron energy to ensure efficient transition to synchrotron acceleration. The betatron bars may then saturate, while synchrotron acceleration maintains the energy increase.

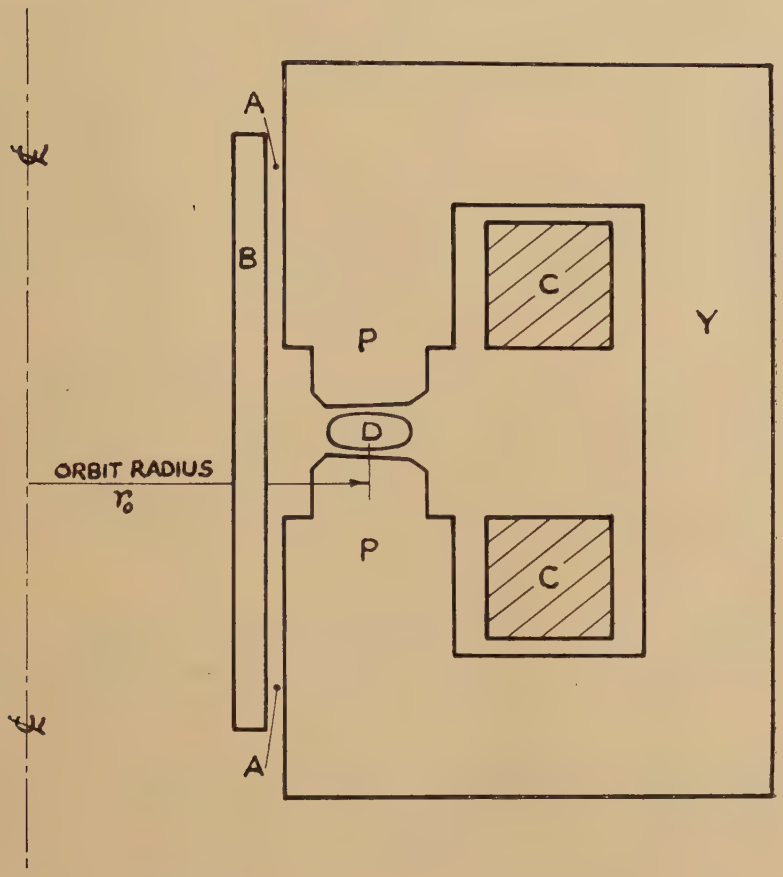
Two fundamental criteria govern satisfactory betatron bar performance. The first is that, to obtain efficient electron injection into the betatron phase, the bars must produce an equilibrium orbit located centrally in the vacuum donut. The second is that, during the betatron phase, the electron orbits must be maintained sufficiently central in the donut to permit efficient transition to synchrotron acceleration when the R.F. accelerating voltage is applied. As Goward (1949) has shown, this second criterion involves marked differences between machines of various energies.

For the sake of definiteness, attention in this paper is focused primarily on synchrotrons producing from about 100 to 400 MeV., and having a general magnet arrangement of the type illustrated in fig. 1. With the exception of §§3 and 4, however, the treatment throughout applies,

* Communicated by Denis Taylor.

without essential modification, to other energy ranges and magnet designs. In §2 it is shown how the first performance criterion mentioned above determines the dimensions of the air gap (or gaps) controlling the betatron flux. The second criterion, namely efficient betatron-synchrotron transition, effectively determines the minimum permissible betatron bar cross section. §§3 and 4 show how the simple transition theory given

Fig. 1.



General arrangement of a synchrotron magnet.

A. Air gaps controlling betatron flux.
B. Betatron flux bar.
C. Main excitation coils.
D. Vacuum donut.
P. Magnet poles.
Y. Magnet return path.

by Goward (1949) may be put into a convenient form for specifying betatron bar performance, treating the problem in full only for synchrotrons of the energy range mentioned above (100 to 400 MeV.). §§5 and 6

then give a method for determining an appropriate betatron bar cross section, employing the D.C. hysteresis loop of the bar laminations, and §7 shows how a moderate amount of "skin effect" in the laminations may be allowed for. No account has been attempted of the effects of cross fluxes in any part of the magnetic circuits, although in practice such fluxes are difficult to avoid entirely.

In Appendix I electrical methods of modifying the betatron-starting flux in a synchrotron are briefly reviewed. In Appendix II, some detailed performance characteristics of the 375 MeV. Glasgow University synchrotron are presented, illustrating various topics discussed in the main paper. A particular attempt is made to estimate overall electron trapping efficiencies more accurately than is possible by Goward's much simplified theory.

§2. Betatron Air Gap Design.

(i) *Fundamental Equilibrium Orbit Equations.*

To establish an equilibrium orbit of radius r_0 in a betatron, the well known Wideröe flux condition must be satisfied :

$$\partial\phi_0/\partial t = 2\pi r_0^2 \partial \mathrm{H}_0/\partial t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

H_0 is the vertical field at radius r_0 in the annular magnet gap. and ϕ_0 is the total flux linking the orbit. Internal leakage fluxes of the synchrotron magnet, together with the flux in the annular gap at radii less than r_0, contribute an appreciable fraction of ϕ_0, and may be expressed as $(1-k_0)2\pi r_0^2 \mathrm{H}_0$, where, for a given magnet, k_0 depends only on r_0. The remainder of ϕ_0 is the flux ϕ_b carried by the betatron bars. Thus ϕ_b and r_0 are related by the equation

$$\partial\phi_b/\partial t = 2\pi r_0^2 k_0 \partial \mathrm{H}_0/\partial t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Assuming that r_0 denotes some fixed radius, a slight adjustment of the bars will change $\partial\phi_b/\partial t$ in relation to $\partial \mathrm{H}_0/\partial t$, and ϕ_b will satisfy an equation

$$\partial\phi_b/\partial t = 2\pi r_0^2 k_0 (1+\eta)\partial \mathrm{H}_0/\partial t. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The equilibrium orbit will now be of radius $r_0+\delta r_0$, where, provided η and $(\delta r_0/r_0)$ are both small,

$$\delta r_0/r_0 = k_0\eta/(1-n). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The field in the annular gap is assumed to follow the usual r^{-n} radial law.

To illustrate the orders of magnitude involved in equation (4), values appropriate to the 375 MeV. Glasgow synchrotron are $n=0{\cdot}7$, $r_0=125$ cm., $k_0=0{\cdot}82$. It follows that a misadjustment in the rate of change of ϕ_b in this machine of only 0·3 per cent involves an error in equilibrium orbit location of 1 cm.

(ii) *Magnetic Circuit Equations.*

To express the Wideröe flux condition in practical terms we must consider the physical layout of the synchrotron magnet, *e. g.* as shown in fig. 1. For simplicity of analysis any actual arrangement of betatron bars may be represented nearly enough by a single laminated bar of

length L, useful cross section A_b, and carrying a uniform flux density B_b, placed in series with an air gap of length g, area A_g in which the flux density is B_g. Thus for continuity we have

$$\phi_b = A_b B_b = A_g B_g. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Let G be the height of the main annular air gap of the synchrotron, at radius r_0, and let H_b denote the magnetic intensity in the betatron bar. Assuming equal return path reluctance for annular and betatron bar fluxes we have

$$g' B_b + L H_b = G H_0 + D, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where $g' = g A_b / A_g$ and the term D represents any D.C. bias which may be applied to the betatron bars.

When skin effect in the bar laminations is negligible, B_b and H_b are simply related by a normal D.C. hysteresis loop taken well into saturation. When skin effect is appreciable, B_b becomes the average flux density in a lamination and H_b denotes the intensity at a lamination surface, and some modification of the hysteresis loop is required to give the correct relation. This problem is not considered further here, since the results of § 2 (iii) are not affected, but a method of constructing a suitably modified loop is given later in § 7.

Differentiating equation (6) partially with respect to time, and using equation (5) gives

$$\partial \phi_b / \partial t = (G/R) \partial H_0 / \partial t, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where R is defined by

$$R = (g/A_g) + (L/sA_b), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

and s is the "differential permeability" $\partial B_b / \partial H_b$. It follows from equation (2) that an equilibrium orbit of radius r_0 will be established when R satisfies the equation

$$G/R = 2\pi r_0^2 k_0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

(iii) *Betatron Air Gap Dimensions.*

To determine the betatron air gap dimensions it is permissible to assume that, in equation (8), $(L/sA_b) \ll g/A_g$ at low flux densities. Equation (9) then approximates to

$$g/A_g = G/(2\pi r_0^2 k_0). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

For economy g and A_g are made as small as possible, the limiting factor being the tolerance on the dimension g. Equations (4) and (7) show that a small error, δg, in g involves an error in equilibrium orbit location given by

$$\delta r_0 / r_0 = -k_0 \delta g / (1-n) g. \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Observations on a 30 MeV. synchrotron (Fry *et al.*, to be published) suggest that, if r_0 denotes the mean radius of the vacuum donut, within 90 per cent of the maximum electron injection into the betatron phase will be obtainable provided δr_0 is less than 0·1 of the radial donut width.

Applying this information, for example, to the 375 MeV. synchrotron, which has $g=1{\cdot}9$ cm. and a donut width of about 12·5 cm., δr_0 must be less than 1·25 cm. From equation (11), g must be correct to 0·007 cm.

Thus equations (10) and (11), plus the requirement for accurate location of the equilibrium orbit at injection, decide the values of g and A_g, and the value of L will generally follow geometrically from these. So far the discussion has involved only a slight restriction on A_b, namely that the approximate equation (10) should be justified. To specify A_b more definitely requires some further specification of betatron bar performance, *i. e.* the rate at which the electron orbits may contract, as a result of bar saturation, while still giving efficient betatron-synchrotron transition.

§3. Betatron-synchrotron Transition.

(i) *Transition Theory.*

The problem of electron trapping at betatron-synchrotron transition has been discussed in several papers on synchrotron theory (*e. g.* Bohm and Foldy 1946), and particular aspects, such as the effect of finite time of rise of the synchrotron R.F. voltage pulse, have been studied (DePackh and Birnbaum 1948; Twiss and Frank 1949). A simple treatment which is sufficiently comprehensive for the purposes of this paper is that given by Goward (1949), who assumes sudden cessation of betatron acceleration and simultaneous application of a steady R.F. voltage. Goward shows that quite different factors determine efficiency of transition in a synchrotron giving say 30 MeV. and one giving a few hundred MeV. To avoid excessive length, the discussion in this section (§3 (i)) is confined to synchrotrons giving from about 100 to 400 MeV., other energy ranges receiving only brief consideration in §3 (ii).

Following Goward's treatment, we assume that electrons are accelerated by betatron action up to a total energy E_s. An R.F. voltage of peak value V is then switched on, trapping a fraction f_1 of the electrons into stable synchrotron orbits. In a high energy machine, Goward shows that V is always large enough to make f_1 approximately unity; *i. e.* nearly all the electrons are trapped. The trapped electrons execute radial oscillations about the equilibrium synchronous orbit (Bohm and Foldy 1946), ranging in amplitude up to a stable limit ρ_s, where ρ_s is a function of V. If V is increased beyond a certain value, V_1, ρ_s exceeds the clearance (ρ_m) between the equilibrium synchronous orbit and the donut wall, and some trapped electrons are consequently lost. V_1 and ρ_m are related by the equation

$$V_1=[(1-n)^2\beta_s^4\pi KE_s/2](\rho_m/r_0)^2, \quad . \quad . \quad . \quad . \quad . \quad (12)$$

where K is given by

$$K=1+n/[(1-n)\beta_s^2]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

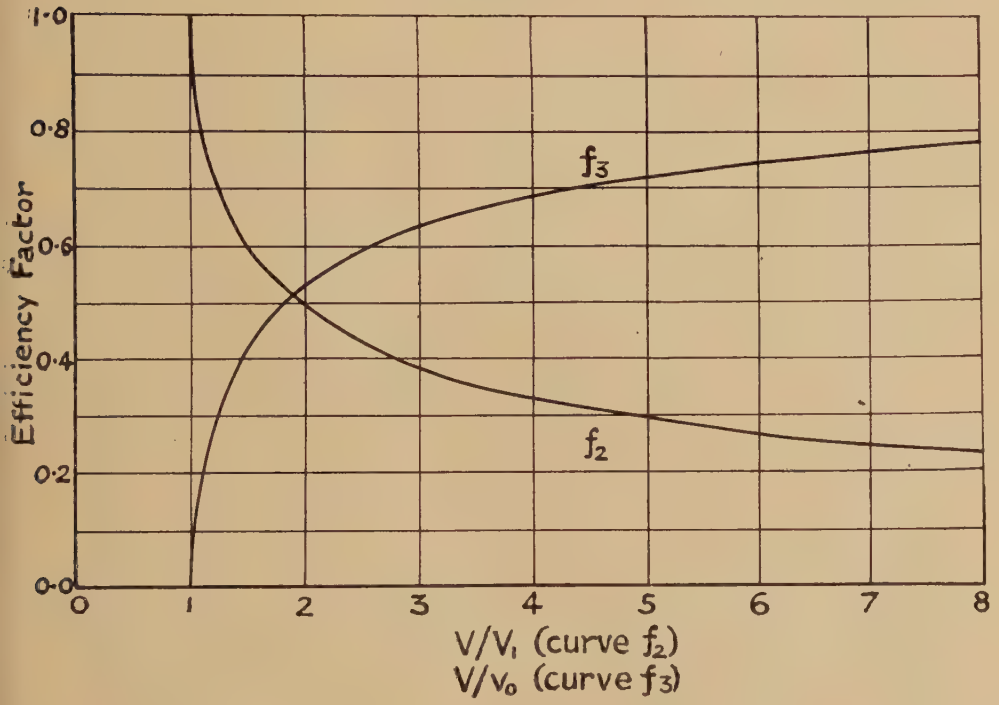
β_s is the electron velocity, at total energy E_s, relative to that of light.

For $V > V_1$ a fraction $(1-f_2)$ of the electrons collide with the donut walls, while f_2 are accelerated further ; f_2 is given by the equation

$$f_2 = (2/\pi)\sin^{-1}(V_1/V)^{\frac{1}{2}}, \qquad \ldots\ldots\ldots \quad (14)$$

shown graphically in fig. 2.

Fig. 2.



Electron trapping efficiency factors.
f_2. Result of scatter to donut walls at transition.
f_3. Result of change of stable R.F. phase during synchrotron acceleration.

As synchrotron acceleration proceeds the betatron flux bars saturate and the accelerating " volts per turn ", v, which must be supplied by the R.F. resonator, rise to a maximum v_0. Assuming that electron radiation losses are not of primary importance in synchrotrons of the energy range considered, v_0 is given by the equation

$$v_0 = 4\pi^2 f r_0 E k_0/(3\times 10^{10}), \qquad \ldots\ldots\ldots \quad (15)$$

where f is the synchrotron magnet excitation frequency in c/s. and E the final (total) electron energy. Thus $V > v_0$ is a necessary condition for synchrotron acceleration and Goward introduces a further efficiency factor f_3, depending on the ratio V/v_0, which takes into account electrons lost during synchrotron acceleration. The factor f_3 is also shown in fig. 2

and is calculated for high energy machines by neglecting phase damping of the synchrotron oscillations (Bohm and Foldy 1946). This damping is, of course, important, but is difficult to include in a sufficiently general manner. To a sufficient approximation, the fraction of original betatron electrons accelerated to peak energy is given by f_2 or f_3, whichever is the least, and hence the most efficient value of V to use is one which makes $f_2=f_3$, *e. g.* for 65 per cent efficiency, putting $f_2=f_3=0{\cdot}65$, fig. 2 gives $(V/v_0)=3{\cdot}3$ and $V/V_1=1{\cdot}35$. Thus we have

$$V_1=2{\cdot}45\,v_0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

By means of an example, it is shown in the next section that an equation such as (16) forms a convenient basis for specifying betatron flux bar performance. It must be emphasized, however, that the equation is based on essentially pessimistic calculations of electron trapping efficiency. Even a simple rectangular R.F. pulse has in practice an appreciable time of rise, and phase oscillation damping will reduce loss of electrons to the walls. The same damping appreciably reduces loss during synchrotron acceleration, and thus both f_2 and f_3 tend to become nearer unity. Moreover it is possible to reduce both kinds of loss by suitable shaping of the R.F. pulse (DePackh and Birnbaum 1948; Twiss and Frank 1949). A relatively small R.F. voltage would suffice for initial trapping of the electrons ($f_1\simeq 1$), after which, to allow for the increasing volts per turn v, the applied R.F. voltage would have to be increased. This latter process in itself gives increased oscillation damping.

(ii) *Flux Bar Performance.*

Taking the 375 MeV. synchrotron as example, the parameters in equation (15) are $f=50$ c/s., $r_0=125$ cm., E=375 MeV., $k_0=0{\cdot}82$. Hence $v_0=2{\cdot}5$ kV., and from equation (16) the condition for 65 per cent efficient overall electron trapping is $V_1=6{\cdot}1$ kV. In equation (12), however, $n=0{\cdot}7$, and assuming $\beta_s\simeq 1$ gives

$$V_1=(3/10^5)E_s\rho_m^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Since $V_1=6{\cdot}1$ kV. we have

$$E_s\rho_m^2=203\times 10^6. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

Equation (18) defines pairs of values of E_s and ρ_m which permit 65 per cent efficient transition, on the basis of the above simple theory. Since the effective donut width is 12·5 cm., if we write $(6{\cdot}25-\rho_m)=\Delta r$, the distance Δr represents the permissible radial displacement of the betatron beam centre from r_0, at transition energy E_s (assuming that transition occurs when betatron beam and stable synchronous orbit coincide). A table of values of E_s and Δr may now be drawn up, as in Table I.

TABLE I.

E_s(MeV.)	5·2	6·2	7·4	9·0	11·3
Δr(cm.)	0	0·5	1·0	1·5	2·0

Thus in addition to establishing an equilibrium orbit of radius r_0, when initial electron injection into the betatron occurs, the betatron bars must satisfy the additional requirement of maintaining the electron orbits constant within the limits of Table I. The method of extending the bar design to meet such a requirement is described in the following sections.

It is convenient at this point to consider briefly how the discussion in this section (§3) must be modified for synchrotrons of other energies. Above 400 MeV., electron radiation loss becomes important, and may determine the maximum R.F. " volts per turn " v_0 (see Appendix II). Moreover R.F. pulse shaping becomes essential to avoid low f_2 values (DePackh and Birnbaum 1948) and the values of E_s and Δr will be governed by a more complex transition theory than that of Goward (1949).

Considering energies below 100 MeV., the values of v_0 and of V become smaller, and loss to the donut walls at transition (f_2) becomes less likely. The value of Δr is not so critical, although Δr must always be less than half the donut width (*i. e.* the stable synchronous orbit and betatron beam must always coincide within the donut limits). The factor f_1, however, which depends on E_s and V, now becomes more important than f_2 or f_3, and the design problem consists essentially of arranging a high enough value of E_s to make $f_1 \simeq 1$, taking into account the value of V available. A full treatment is given in Goward's paper.

§4. Betatron Flux Bar Cross Section.

(i) *Instantaneous Circle Orbit Equations.*

Neglecting free oscillations, individual electrons in a betatron beam move on so-called " instantaneous circle " orbits (Kerst and Serber 1941) which, at any time, have a range of radii corresponding to the energy spread of the electrons. It is convenient, for the following discussion, to consider the electron motion in terms of these instantaneous circle orbits, and to use the instantaneous circle orbit equation (Amaldi and Ferretti 1946):

$$\phi_r = 2\pi r^2 H_r + C. \qquad (19)$$

H_r is the vertical field at radius r and ϕ_r is the total flux linking the circle of radius r. If C is a constant, the equation defines the variation of r for any electron throughout its betatron motion.

For design purposes it is sufficient to consider an electron injected from the electron gun so that, initially, $r = r_0$. (This electron roughly defines the mean orbit radius of the whole electron beam.) At injection, equation (19) for this electron takes the form

$$\phi_0 = 2\pi r_0^2 H_0 + C. \qquad (20)$$

After some betatron acceleration, r will in general have changed to some value $r_0 + \Delta r$, where we assume $(\Delta r / r_0)$ is small. Corresponding to this, the flux through the circle r_0 will then be given by

$$\phi_0 = 2\pi r_0^2 H_0 + C + \Delta\phi_0, \qquad (21)$$

where $(\Delta\phi_0/\phi_0)$ is also small. $\Delta\phi_0$ and Δr are related by the equation

$$\Delta\phi_0/(2\pi r_0^2 H_0)=(1-n)(\Delta r/r_0). \quad . \quad . \quad . \quad . \quad . \quad (22)$$

(ii) *Determination of Betatron Bar Cross Section.*

From §2 the betatron air gap dimension and bar length L (as defined in §2) are already determined. For the present purpose, H_0 is nearly enough zero when injection into the betatron occurs. Thus when $H_0=0$, let the variables in equation (6) have the values $B_b=B_1$, $H_b=H_1$. Then for any value of H_0 we have

$$B_b-B_1=(G/g')H_0-(L/g')(H_b-H_1). \quad . \quad . \quad . \quad . \quad (23)$$

Equation (23) gives the total change in B_b corresponding to an increase in field at r_0 from zero to H_0.

A theoretically ideal betatron bar would have constant differential permeability s, and would maintain the electron considered in §4(i) at a constant orbit radius $r=r_0$. Let the parameter R in equation (7) have the value R_1 when $B_b=B_1$ (and $s=s_1$). Then, for an ideal bar, R would be constant at R_1, and equation (7) gives

$$B_b-B_1=GH_0/(A_bR_1). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

Comparison of (23) and (24) gives the amount, ΔB_b, by which the actual bar density differs from the ideal value, as a result of variation of s and R. Assuming all non-linear effects to be confined to the betatron bar flux, we may identify $A_b\Delta B_b$ with the $\Delta\phi_0$ of equation (22), and obtain

$$(R_1/L)\Delta\phi_0=(B_b-B_1)(1/s_1)-(H_b-H_1). \quad . \quad . \quad . \quad (25)$$

If we neglect skin effect for the moment, the right hand side of equation (25) has a simple graphical meaning in terms of the D.C hysteresis loop as shown in fig. 3, and may conveniently be denoted by $(-\Delta H_b)$. Using equations (9) and (22), we have

$$-\Delta H_b=[(1-n)G/Lr_0k_0]H_0\Delta r. \quad . \quad . \quad . \quad . \quad . \quad (26)$$

Further, using the well known relation

$$E_s\beta_s=300r_0H_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

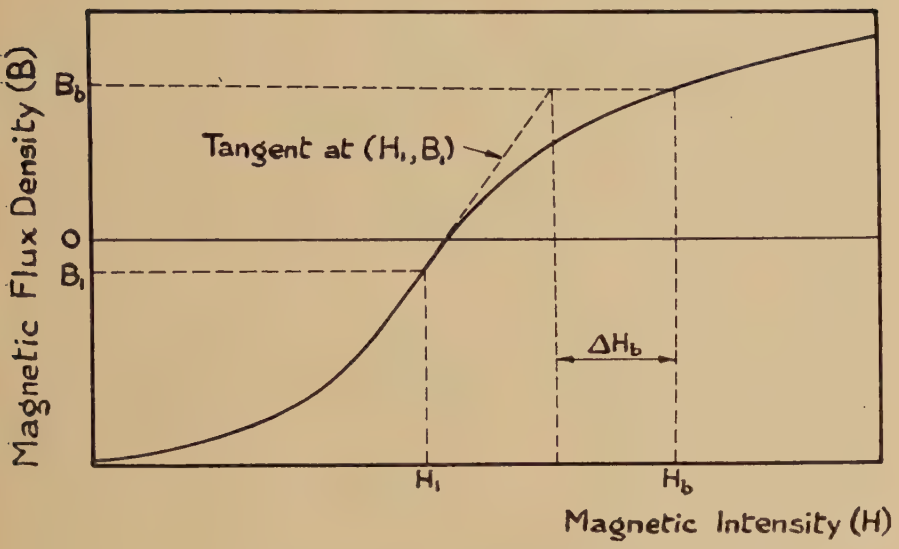
equation (26) may alternatively be written

$$-\Delta H_b=[(1-n)G/(300\,Lr_0^2k_0)]E_s\beta_s\Delta r. \quad . \quad . \quad . \quad . \quad (28)$$

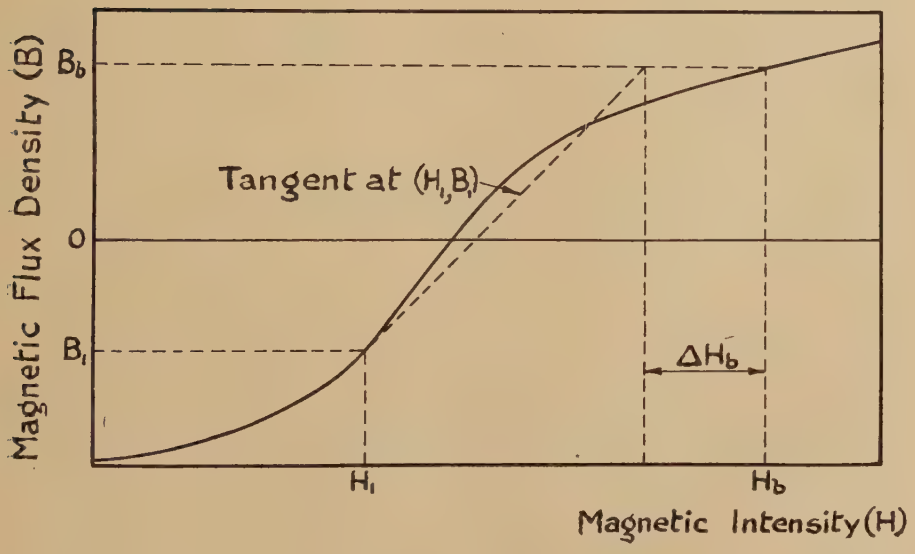
In conjunction with a pair of values of E_s and Δr, specified for example as in §3, equation (28) effectively determines a betatron bar cross section. The derived value of ΔH_b specifies pairs of values of B_b and B_1 from the hysteresis loop of the bar material, as in fig. 3. From equations (9), (24) and (27), A_b is then given in terms of such a pair of values by the equation

$$A_b=2\pi r_0E_s\beta_sk_0/300(B_b-B_1). \quad . \quad . \quad . \quad . \quad . \quad (29)$$

Fig. 3.



(a) Little (or no) Negative Flux Bias.



(b) Large Negative Flux Bias.

Hysteresis loop construction for determining changes in electron orbit radius.

In the following section, the application of the above procedure is illustrated by a specific example.

§5. Example Design : 375 MeV. Synchrotron.

(i) *No Applied D.C. Flux Bias.*

Taking again the example of the 375 MeV. Glasgow synchrotron, the known parameters in equation (28) are $n=0{\cdot}7$, G=10 cm., L=100 cm., $r_0=125$ cm. and $k_0=0{\cdot}82$. If at least 65 per cent efficient transition is taken as a design criterion, Table I. gives appropriate values of E_s and Δr. For example, the values $E_s=9{\cdot}0$ MeV., $\Delta r=1{\cdot}5$ cm., inserted in equation (28), give $-\Delta H_b=0{\cdot}105$. Drawing a tangent at the point $B_1=0$ to the hysteresis loop of the transformer steel laminations to be used yields the value $B_b=4600$ gauss, in the manner indicated in fig. 3 (*a*). From equation (29), $A_b=4200$ sq. cm.

Other values of E_s and Δr from Table I. give cross sections greater than 4200 sq. cm. Hence we conclude that if no D.C. bias is to be used, and skin effect in the laminations is negligible, a cross section of 4200 sq. cm. is the minimum which will ensure at least 65 per cent transition efficiency.

(ii) *Application of D.C. Bias.*

An auxiliary D.C. winding on the betatron bars may be used to give B_1 some value other than zero, and to effect a saving in bar cross section. The graphical construction may now take the form of either fig. 3 (*a*) or 3 (*b*), according to the amount of bias used. 2000 gauss negative bias, for example, gives a fig. 3 (*a*) type of construction, and the same design procedure as in §5 (i) yields a cross section of 3200 sq. cm. A greater bias gives the type of construction shown in fig. 3 (*b*). Here, as is readily seen from equation (26) or (28), the electron orbit, initially of radius r_0, will first expand to some extent, and then contract through r_0 as tangent and hysteresis loop intersect. The design procedure is unaltered, and taking B_1 as 4000 gauss negative gives a cross section of 2000 sq. cm., less than half that determined for zero bias.

Equation (26) may be used to check that the initial orbit expansion under fig. 3 (*b*) conditions is not excessive ; *e.g.* in the above 4000 gauss bias example, instantaneous circle orbit expansions from 125 cm. at injection to a maximum of 126 cm. at about 3·6 MeV. electron (total) energy will occur.

By the above methods a betatron bar cross section can be determined to give any specific degree of electron orbit contraction. One single value of cross section is obtained if no flux bias is to be used, but a variety of combinations of cross section and bias will give essentially the same results. So far the question of skin effect in the bar laminations has been neglected and we now proceed to determine whether practical lamination thicknesses involve any serious modifications in betatron bar performance.

§6. Skin Effect.

(i) *Modification of Hysteresis Loop.*

The use of practical lamination thicknesses in betatron bar construction will in general involve appreciable skin effect, and the variables B_b and H_b

used in §§4 and 5, should strictly be defined respectively as the average flux density in a betatron bar lamination and the intensity at a lamination surface. Some modified hysteresis loop is therefore required to relate B_b and H_b if the graphical method of §4 is to yield a correct result.

Consider a lamination extending in thickness between the limits $-\xi/2 \leqslant x \leqslant \xi/2$. The flux density, B, and intensity, H, at any point in the lamination (directed say along the y axis) are considered to be functions of x and time, t, only, and are related by the well known equation

$$\partial^2 H/\partial x^2 = (4\pi/10^9\rho)\partial B/\partial t, \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

where ρ is the electrical resistivity of the material. Solutions of the equation when B is a linear or sinusoidal function of t are also well known, but to a first approximation the particular function of time is immaterial, and it is found that $\partial B/\partial t$ may be treated as a constant with respect to x. Equation (30) is then simply integrable, and it follows that the distribution of H is of the form $(a+bx^2)$. If H_b and H_c are the values of H when $x = \pm\xi/2$ and 0 respectively, then

$$H_b - H_c = (\pi/10^9 . 2\rho)\xi^2 \partial B_b/\partial t, \quad . \quad . \quad . \quad . \quad . \quad (31)$$

where B_b is the value of B averaged over x.

Assuming that the synchrotron magnet excitation is derived from a resonant circuit of frequency f, the magnetic field H_0 will not vary sinusoidally, since saturable betatron bars introduce non-linearity into the circuit. If the ratios of magnet inductances with unsaturated and saturated bars is q, then during the initial betatron phase

$$\partial H_0/\partial t = 2\pi f H_m/q, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

where H_m is the peak value of H_0. Using equations (7), (9), (27) and (31) we obtain

$$H_b - H_c = (2{\cdot}07/10^{10}) r_0 k_0 f E \xi^2/q A_b \rho, \quad . \quad . \quad . \quad . \quad (33)$$

where E is the peak electron energy (eV.).

Corresponding to the range of variation of H given by equation (33), the flux density B will vary from B_s at the lamination surface to B_c at its centre, where (H_b, B_s) and (H_c, B_c) are points on the normal D.C. hysteresis loop. If we approximate to this range of the hysteresis loop by a straight line, the distribution of B will also be of the form $(a+bx^2)$, and by simple integration

$$B_b - B_c = (B_s - B_c)/3. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

As indicated in fig. 4 this approximate result gives a graphical method for deriving from the D.C. hysteresis loop an (H_b, B_b) curve representing true operating conditions.

(ii) *Example*: 375 *MeV. Synchrotron.*

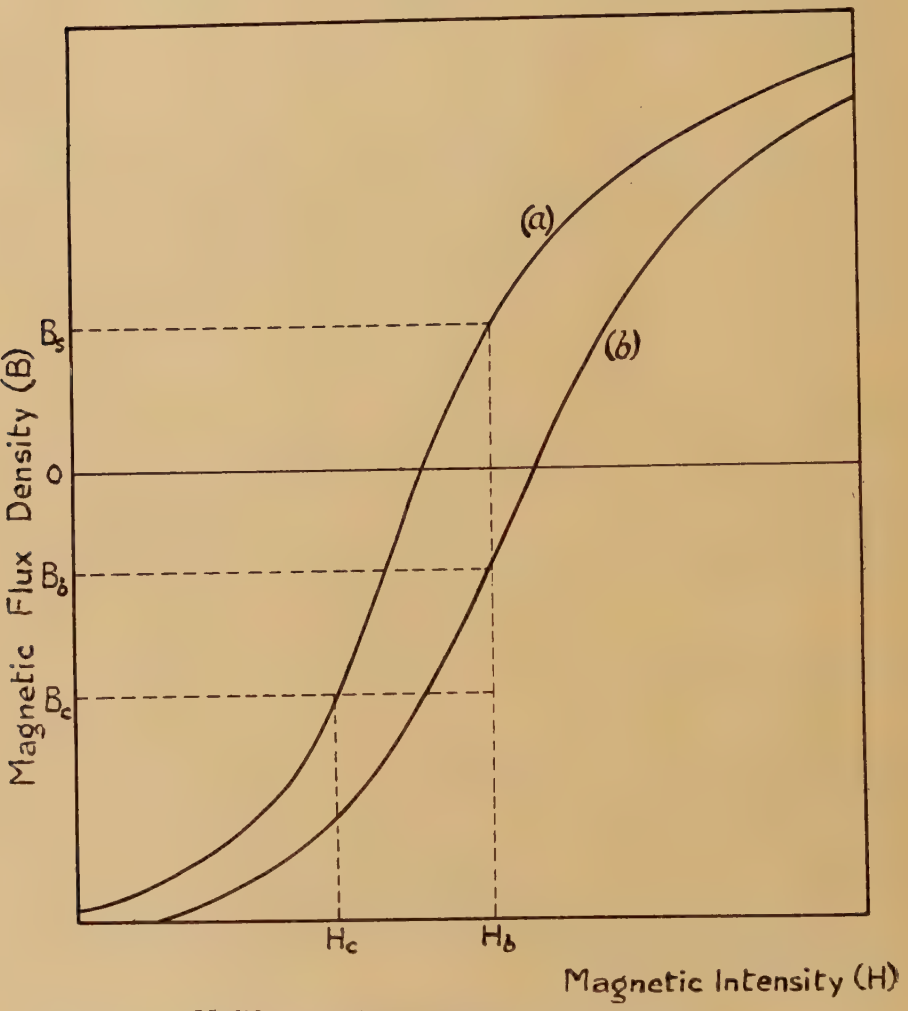
Again taking a specific example, the 375 MeV. synchrotron, the proposed bar material consists of 0·007 in. laminations of resistivity

50 microhms per cm. cube. With $r_0=125$ cm., $k_0=0{\cdot}82$, $f=50$ c/s. and $q=3$, equation (33) gives

$$H_b-H_c=830/A_b, \quad \ldots\ldots\ldots \quad (35)$$

if A_b is expressed in sq. cm. By neglecting skin effect, a cross section of 4200 sq. cm. was determined as suitable for use with no D.C. bias.

Fig. 4.



Modification of hysteresis loop by skin effect.
(*a*) D.C. hysteresis loop,
(*b*) Modified loop $B_b-B_c=(B_s-B_c)/3$.

For this cross section $(H_b-H_c)=0{\cdot}2$, which is small enough for the approximate methods of the last section to be applied. Deriving the modified loop, as in fig. 4, and applying the construction of fig. 3 (*a*), with $B_1=0$, indicates orbit contractions slightly less than those predicted by the

unmodified loop. Neglecting skin effect resulted in an over estimation of the cross section required by only a few per cent.

Similar calculations may be made on the assumption of A_b values determined in §5(ii). The practical result is to predict that, for a given cross section, rather less bias would be required than that specified in §5(ii). For example a 2000 sq. cm. cross section would require about 3000 rather than 4000 gauss negative bias. The detailed variations in electron orbit radius are of course affected both by the change of bias and the modified loop shape, and rechecking for excessive initial expansion as in §5(ii), shows expansion up to 127 cm. at 5 MeV. total electron energy.

It is reasonable to conclude that in general, provided equation (33) indicates no excessive skin effect, use of a normal D.C. hysteresis loop is sufficiently accurate for initial design purposes. A final design should, however, be checked by the appropriate modified loop.

§7. Conclusion.

The essential steps in the magnetic design of betatron bars for an electron synchrotron may now be summarized as follows. Betatron air gap dimensions are determined from equations (10) and (11), and the flux bar length generally follows simply from geometrical considerations. The methods of §3 define the rate of radial shrinkage of the electron beam orbit which may be permitted if efficient betatron-synchrotron transition is to be ensured. Equation (28) essentially expresses orbit shrinkage rate in terms of non-linearity of the betatron bar material, and a graphical construction on the hysteresis loop, as shown in fig. 3, together with equation (29) gives the appropriate cross sectional area of bar. As a check on the validity of neglecting skin effects equation (33) should be evaluated, and a final design should be checked by proceeding to construct a modified hysteresis loop by the method of §6(i) and fig. 4.

Acknowledgments.

Acknowledgment is made to many helpful suggestions made to the author at Metropolitan Vickers Electrical Company, by members of the Glasgow Synchrotron Committee. This paper was written as part of the accelerator programme of the Electronics Division of the Atomic Energy Research Establishment, and is published with the permission of the Director.

References.

Amaldi, E., and Ferretti, B., 1946, *Rev. Sci. Instrum.*, **17**, 389.
Blewett, J. P., 1945, *Phys. Rev.*, **69**, 87.
Bohm, D., and Foldy, L., 1946, *Phys. Rev.*, **70**, 249.
DePackh, D. C., and Birnbaum, M., 1948, *J. Appl. Phys.*, **19**, 795.
Goward, F. K., 1949, *Proc. Phys. Soc.* A, **62**, 617.
Kerst, D. W., and Serber, R., 1941, *Phys. Rev.*, **50**, 53.
Kerst, D. W., 1945, *Phys. Rev.*, **69**, 233.
McMillan, E. M., 1945, *Phys. Rev.*, **68**, 144.
Twiss, R. Q., and Frank, N. H., 1949, *Rev. Sci. Instrum.*, **20**, 1.

APPENDIX I.

Electrical Control and Correction of Betatron Flux.

The effect of applying D.C. bias to betatron-starting bars has already been considered in the main paper, and it is shown that considerable improvement in performance may be obtained. Other methods of counteracting non-linearity in the betatron bars by auxiliary electrical circuits are available, and the techniques involved are being developed at Metropolitan-Vickers Electrical Co., Ltd., in association with the 375 MeV. synchrotron magnet design. However, a brief review may be appropriate to this paper.

One possible general technique is suggested by Kerst's (1945) flux forcing circuit, which consists of an auxiliary winding around main poles and betatron bars, automatically ensuring the correct relation between annular and central fluxes. In application to a synchrotron the circuit must be "switched off" after betatron-synchrotron transition has occurred, to avoid excessive currents. Mechanical or electronic switching are possible solutions, if care is taken over transient effects. A more elegant modification is an electronic circuit in which, for example, any unbalanced voltage from a Kerst type winding is fed into an amplifier which in turn feeds a correcting winding on the betatron bars. Switching should then be relatively simple.

Electrical control may also be used to give fine adjustment of the equilibrium orbit radius, an essential operational facility. Mechanical adjustment of the betatron air gap alone may be neither convenient nor sufficiently sensitive. (See for example, the air gap tolerance figures of §2(iii)). An auxiliary bar winding connected across a variable inductance conveniently achieves the desired result.

APPENDIX II

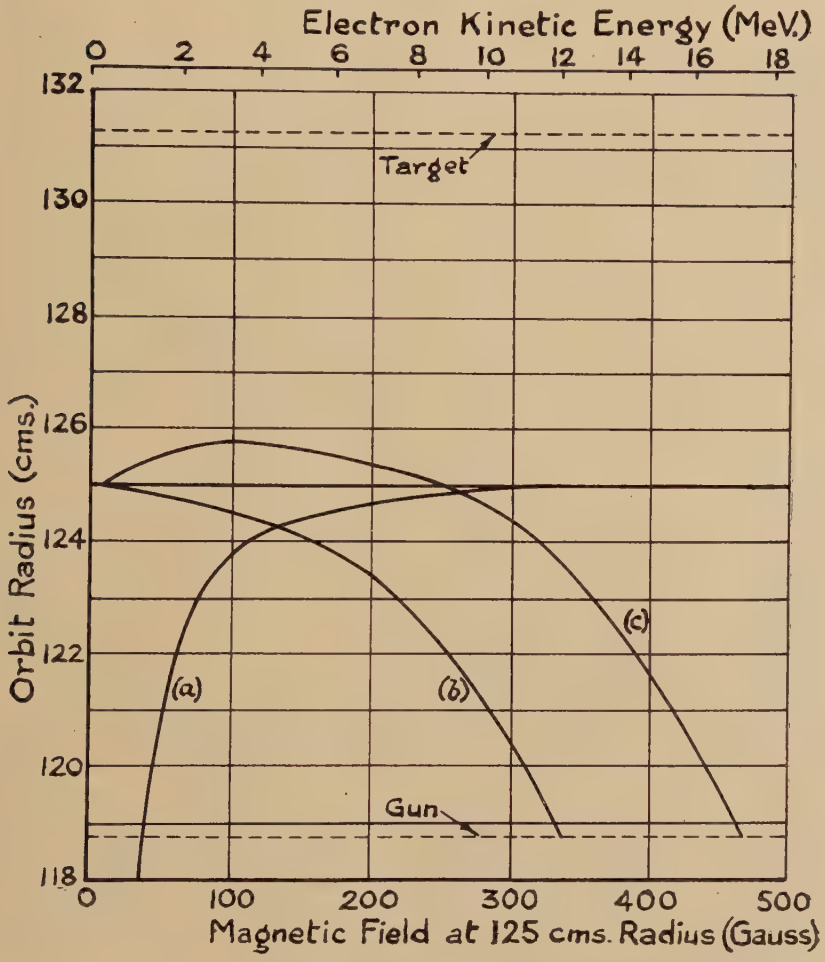
375 MeV. Synchrotron Performance Characteristics.

In the main paper the design methods discussed were illustrated at various points by calculations based on the 375 MeV. synchrotron, and in particular it was determined (§§3 (ii) and 5 (i)) that a betatron bar of 100 cm. effective length and 4200 sq. cm. cross section, of the specified material and used without D.C. bias, would make at least 65 per cent betatron-synchrotron transition possible. In the specification finally adopted for the machine, the effective flux bar dimensions are 100 cm. length and 3300 sq. cm. cross section, and using this specification it was considered worth while to calculate the expected performance in some detail. Some of the more instructive results of the calculations are presented below. Since the machine will initially be operated at 300 MeV., some results for this energy are included.

Using a hysteresis loop corrected for skin effect by the method of §6, the instantaneous circle motion of an electron, initially injected on to the equilibrium orbit, was calculated for various flux bar D.C. bias values, by

means of equation (26). The difference in skin effect for 300 and 375 MeV. operation is too small to affect the results appreciably. Fig. 5 shows the calculated orbits, and a superposed curve shows the synchronous orbit radius expanding towards r_0 as the electron velocity approaches that of

Fig. 5.



Variation of orbit radii.
375 MeV. synchrotron.

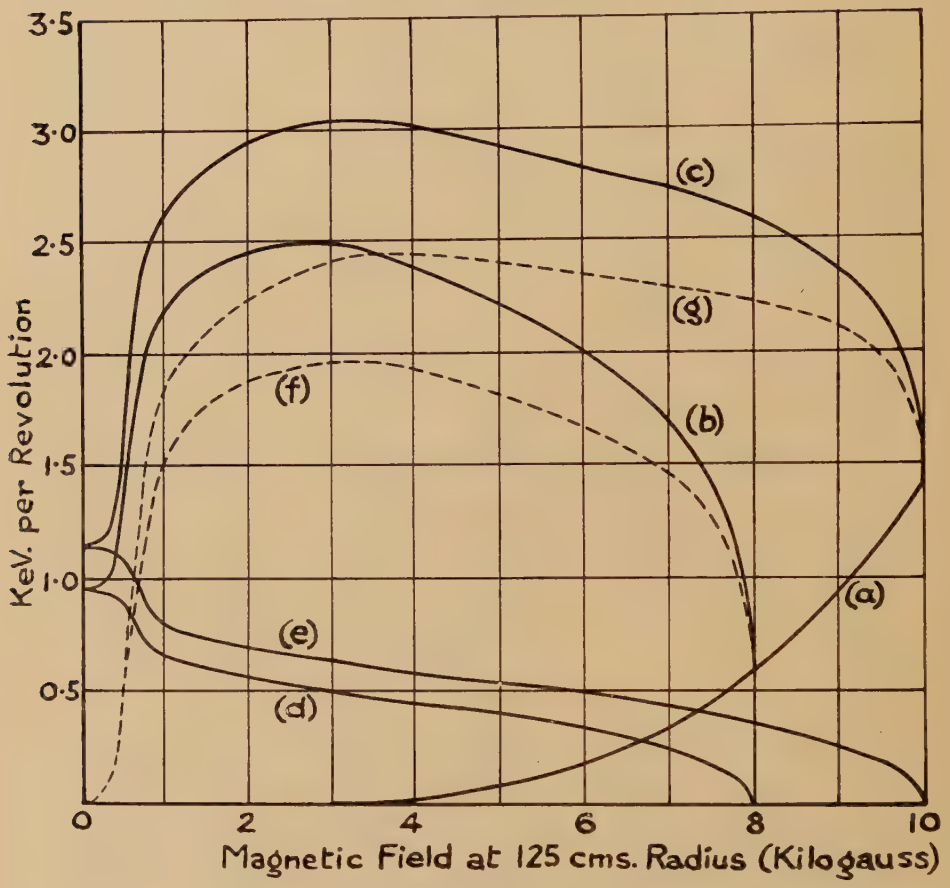
(*a*) Synchronous orbit.
(*b*) Betatron electron orbit. Zero betatron bar flux bias.
(*c*) Betatron electron orbit. 2000 gauss betatron bar flux bias.

light. Betatron-synchrotron transition takes place at the intersection of this curve with a betatron orbit curve, and it would appear that transition conditions may be materially improved by flux biasing or other electrical correction methods. Further adjustment may be acheived by alteration

of the resonator frequency, making the synchronous radius tend to some value other than 125 cm.

The betatron output energy of the machine may also be deduced from fig. 5 ; *e. g.* with no flux biasing, electrons would strike the internal gun at about 12 MeV. energy. (An external target and internal gun are to be used in this machine.)

Fig. 6.



Volts per revolution diagram.
375 MeV. synchrotron.

(*a*) Incoherent radiation loss.
(*b*), (*c*) Total volts per revolution required for acceleration to (*b*) 300 MeV., (*c*) 375 MeV.
(*d*), (*e*) Volts per revolution supplied by betatron effect, (*d*) 300 MeV., (*e*) 375 MeV.
(*f*), (*g*) Volts per revolution derived from R.F. cavity, (*f*) 300 MeV., (*g*) 375 MeV.

To maintain a constant orbit radius the electron energy must increase by $(2\pi r_0^2/10^8)\partial H_0/\partial t$ electron volts per revolution. Incoherent radiation of energy by the electrons (MacMillan 1945) is equivalent to an energy loss (also in eV. per revolution) given by the expression $(6/10^7 r_0)(E_t\, m_0c^2)^4\beta^3$,

where E_t is the total electron energy and m_0c^2 its rest energy. Some coherent radiation loss also occurs, but assuming that the total charge associated with the electron beam will be about 1/100 (or less) of the theoretical maximum value 0·05 microcoulomb (Blewett 1945), this loss is likely to be comparatively small. Thus betatron effect and the R.F. voltage must supply to the electrons an energy represented by the sum of the two above expressions.

Fig. 6 is an "eV. per revolution diagram" showing the relative magnitudes of the above effects. The R.F. volts per turn (v) rises relatively slowly (adiabatically) to the maximum value v_0. Synchrotron phase oscillation damping will considerably reduce electron loss during this period, giving more efficient overall electron trapping than that indicated by Goward's (1949) methods. It is thus of interest to attempt an approximate estimate of the overall efficiency, taking some account of phase damping, but continuing to assume a simple rectangular R.F. voltage wave form.

Bohm and Foldy (1946) and several other authors have discussed the phase oscillations of electrons in a synchrotron about the stable R.F. phase ψ, where ψ is given by

$$\sin\psi = v/\mathrm{V}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

When ψ is zero, stable oscillations are possible up to a full 180° amplitude, but as ψ increases from zero, the permissible amplitude decreases and, moreover, the oscillations become assymetric about ψ. Thus, when ψ is 30° the maximum stable oscillation is from −40° to +150°, which is −70° to +120° about ψ. In order to make the following trapping efficiency calculations pessimistic rather than optimistic, we consider only the smaller phase amplitude (such as 70° when ψ is 30°), and denote this by θ_m. At all times during synchrotron acceleration the electron bunch must contain a range of smaller phase amplitudes from zero up to θ, where $\theta \leqslant \theta_m$.

At transition energy, v and ψ are approximately zero and θ_m is 180°. If some loss to the walls occurs at transition, then from the definition of f_2 in equation (14), the remaining electrons have phase amplitudes up to only $(180° f_2)$. After some synchrotron acceleration has taken place, and ψ is no longer zero, all electrons which tend to have phase amplitudes greater than θ_m are considered lost. Thus, neglecting damping for the moment, of the electrons remaining in the beam immediately after transition a fraction u will now remain, where (if θ_m is expressed in degrees)

$$u = \theta_m/180 f_2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

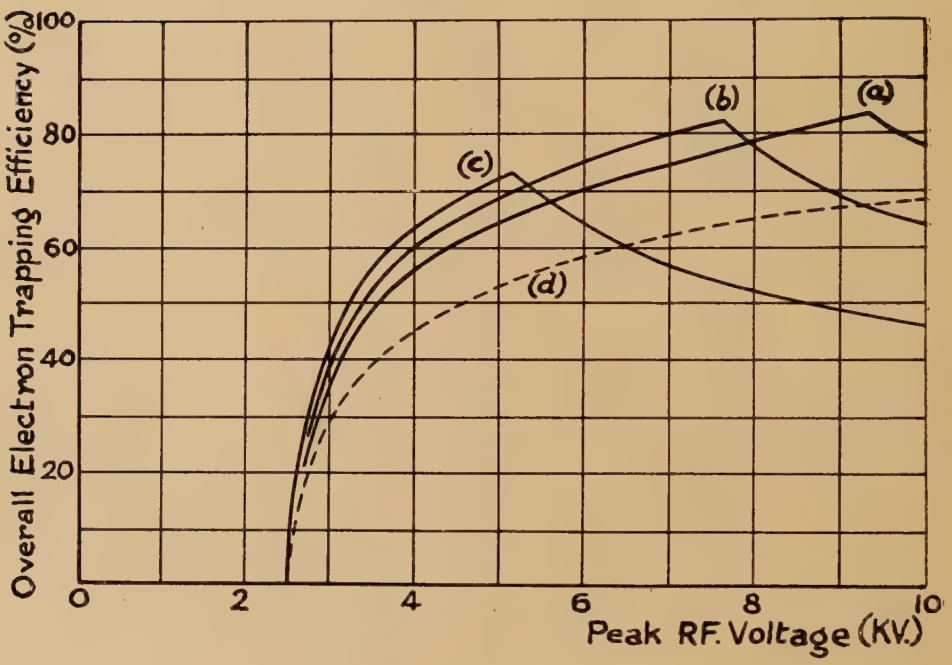
This is essentially the calculation used by Goward (1949).

It has been shown (*e. g.* Bohm and Foldy 1946) that synchrotron phase oscillations of small amplitude are damped in proportion to $(E_t \cos\psi)^{\frac{1}{2}}$. The phase oscillation equation can only be solved for all amplitudes when

ψ is zero (*e. g.* Twiss and Frank 1949), and, from the form of the resulting equation, the rate of damping appears to increase with the amplitude, approximating to $E_i^{\frac{1}{2}}$ over a wide amplitude range. Thus if the small amplitude damping law is assumed to hold for all amplitudes and ψ values, the tendency again should be to a pessimistic rather than optimistic result. Equation (2) is accordingly modified to the form.

$$u=(\theta_m/180f_2)(E_i \cos \psi/E_s)^{\frac{1}{2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Fig. 7.



Overall electron trapping efficiency.
375 MeV. synchrotron.
(*a*) Transition at 8·0 MeV. total electron energy.
(*b*) Transition at 6·5 MeV. total electron energy.
(*c*) Transition at 5·0 MeV. total electron energy.
(At 0·75 cm. off the donut centre).
(*d*) Efficiency factor f_3 (neglecting phase oscillation damping).

Fig. 7 shows overall trapping efficiency curves, determined by the data of fig. 6 and equation (3). Curves (*a*) and (*b*) assume transition exactly at the donut centre, while curve (*c*) represents the uncorrected transition of fig. 5, for which the betatron beam is displaced by 0·75 cm. from the donut centre. For convenience of comparison curve (*d*) has been added, representing the efficiency factor f_3 of fig. 2 and showing the effect of neglecting electron oscillation damping.

As in Goward's calculations, the rising part of an efficiency curve represents conditions under which loss of electrons during the synchrotron phase is predominant, and the falling part is determined by the loss at transition. It appears that an R.F. voltage of about 5 kV. peak will ensure 70 per cent overall trapping efficiency, without flux correction, and higher efficiencies are possible under adjusted conditions. Moreover, the effect of finite R.F. switch on time, which has been neglected, will be to give additional efficiency.

IV. *On the Interaction of Colloidal Particles.*—III. *Some Theorems Concerning the Free Energy of Double Layers.*

By S. Levine, Ph.D.,
Birkbeck College Research Laboratory*.

[Received July 26, 1949.]

Summary.

Making use of the Poisson-Boltzmann equation, a number of theorems relating to the free energy of the electric double layers of colloidal particles in dilute sols are developed. A general proof is given for the equivalence of two methods for determining this free energy, viz. by (i) using the fictitious process of charging the excess ions in the double layers subject to the condition of constant surface density of ions and (ii) imagining that the particles are charged by bringing ions from the interior of the dispersion medium to the surface. In the case of two infinite parallel plates, the difficulty that arises due to the infinity of the surface potential, when all the ions are supposed completely discharged, is satisfactorily explained.

The present work is an extension of some of the results obtained by Verwey and Overbeek in their recent book on the theory of the stability of lyophobic colloids. Our theorem of the equivalence of the two methods of evaluating the free energy is more general than a similar theorem proved by Casimir in this book. We are able to derive the Casimir theorem by introducing special conditions.

---

§ 1. Introduction.

In their recent book, Verwey and Overbeek (1948) develop two methods of evaluating the free energy associated with the double layers of charged colloidal particles, namely by (i) charging the excess ions in the diffuse part of the double layers under the condition of constant surface potential and (ii) using a generalized Lippmann equation. A general proof of the equivalence of these two methods is given in the book by Casimir (1948).

---

* Communicated by the Author.

In this paper we derive a corresponding theorem, where the distribution of surface ions, and not the surface potential, is kept fixed, by means of an argument similar to that used by Casimir. Our equivalence theorem can be derived without the use of the Poisson-Boltzmann equation (Levine 1949b) but the proof given here does depend on this equation (or rather a generalized version of it) and, in addition, only applies to dilute sols. Our theorem is more general than that proved by Casimir, since it is valid if the surface distribution of ions is arbitrary (but unchanged during the charging process) and, furthermore, Casimir's theorem can be deduced by assuming that the surface potential be constant during charging.

Verwey and Overbeek apparently did not investigate our method of charging the ions in any detail because of the anomaly of an infinite potential (for the discharged state) in the case of two infinite parallel plates. This difficulty was also encountered by the author and, in fact, lead to an erroneous conclusion (Levine 1946) concerning the nature of the forces between the plates in dilute sols (an attraction was obtained at large separations, whereas the correct result is a repulsion). However, in §3, we are able to overcome this obstacle in a satisfactory manner.

§2. Energy of Double Layer of Single Particle.

For the sake of simplicity we begin with a single colloidal particle of any shape immersed in a large volume of electrolyte. Let the surface charge arise from the adsorption of ions of type 1, charge e_1: ν is the density of such ions at position $\mathbf{s}$ on the surface. If n is the total number of adsorbed ions, then $n=\int_S \nu\, dS$, integrating over the surface (area) S of the particle. We imagine that all the ions in the system, including the adsorbed ones, are subject to the familiar Debye-Huckel charging process. Assuming that all the ions are being charged at the same rate, the fraction λ will denote the stage in this fictitious process. Our problem is to determine the free energy of the double layer at any stage $\lambda' \leqslant 1$. Now in an earlier paper (Levine 1946), we have already remarked that our method of determining this free energy is based on keeping the surface distribution of ions fixed when increasing λ from 0 to λ'. Thus ν does not depend on any intermediate value $\lambda < \lambda'$. It is also assumed that the distribution ν can be chosen arbitrarily, *i.e.* we treat it as a thermodynamic variable. There is really no loss in generality if we assume that ν has the form

$$\nu = n\nu_0(\lambda', \mathbf{s}), \qquad \int_S \nu_0(\lambda', \mathbf{s})\, dS = 1, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where $\nu_0(\lambda', \mathbf{s})$ is subject to the normalizing condition, but it is otherwise an arbitrary function of λ' (which represents here the completion of the charging process) and of the surface position $\mathbf{s}$. In this way n takes on the simple rôle of a " magnitude " factor.

The density of ions of species $i (i = 1, 2, \ldots, s)$ in the bulk of the solution is assumed constant, independent of n, and will be denoted by n_i^0.

The average potential at position **r** in the diffuse layer of the particle is ψ, which is taken to be the same for the different types of ions. Casimir (1948) writes the Poisson equation for ψ in the form

$$\nabla^2\psi = -4\pi\rho/\mathrm{D} = \lambda f(\lambda\psi), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where ρ is the average charge density in the diffuse layer, D is the dielectric constant of the dispersion medium and f is an arbitrary function. This becomes the Poisson-Boltzmann equation if we put

$$f(\lambda\psi) = -\frac{4\pi}{\mathrm{D}} \sum_{i=1}^{s} n_i^0 e_i \exp(-\lambda e_i\psi/k\mathrm{T}), \quad . \quad . \quad . \quad . \quad (3)$$

where e_i is the (full) charge of an i ion, k is Boltzmann's constant and T is the temperature. A different form for $f(\lambda\psi)$ has been chosen by Grimley and Mott (1947) when they take into consideration the hydration shell about each ion. The general conclusions in this paper do not depend on the particular form of the function $f(\lambda\psi)$. The average electrostatic energy associated with the excess charge in the double layer at stage λ is

$$\mathrm{E}(\lambda, n\,;\ \lambda') = \tfrac{1}{2}\lambda e_1 \int_{\mathrm{S}} \psi\nu \, d\mathrm{S} + \tfrac{1}{2}\int_{\mathrm{V}} \psi\rho \, dv. \quad . \quad . \quad . \quad . \quad (4)$$

The second integral in (4) is taken over the volume V of the outer diffuse layer, which may be assumed to extend to infinity, where the potential ψ is zero. Then the corresponding electrical free energy at stage λ' is given by (see Levine 1948)

$$\mathrm{F}_e(\lambda', n) = 2\int_0^{\lambda'} \frac{1}{\lambda}\mathrm{E}(\lambda, n\,;\ \lambda')\, d\lambda. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The quantities ψ and ρ depend on λ, λ', ν (and hence n) and the coordinates **r** or **s** but to simplify the notation, we shall not indicate this explicitly. It will be clear from the text whether we are referring to the potential at position **r** in the outer layer or at position **s** on the surface.

It is convenient to evaluate the quantity

$$\frac{\partial \mathrm{F}_e(\lambda', n)}{\partial n} = 2\int_0^{\lambda'} \frac{1}{\lambda}\frac{\partial \mathrm{E}(\lambda, n\,;\ \lambda')}{\partial n}\, d\lambda. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Since we keep ν and hence n fixed when integrating with respect to λ, n and λ are considered independent of one another, and we can differentiate under the integral sign with respect to n. The partial derivatives of the quantities ψ and ν with respect to λ or n will be denoted by a subscript; we note that $\nu_\lambda = 0$. Substituting (2) for ρ into (4), we readily obtain

$$\frac{\partial \mathrm{E}(\lambda, n\,;\ \lambda')}{\partial n} = \tfrac{1}{2}\lambda e_1 \int_{\mathrm{S}} (\psi_n\nu + \psi\nu_n)\, d\mathrm{S} - \frac{\mathrm{D}\lambda}{8\pi}\int_{\mathrm{V}} (\psi_n f + \lambda\psi\psi_n f')\, dv, \quad . \quad (7)$$

where f' if the derivative of f with respect to the argument $\lambda\psi$. Now from (2) we have the relations

$$\nabla^2\psi_\lambda = f + \lambda(\psi + \lambda\psi_\lambda)f', \quad \nabla^2\psi_n = \lambda^2\psi_n f'. \quad . \quad . \quad . \quad . \quad (8)$$

Then use of (2) and (8) and application of Green's theorem yields the following formulæ for the volume integral on the right of (7)

$$\frac{D\lambda}{8\pi}\int_V(\psi_\lambda\nabla^2\psi_n-\psi_n\nabla^2\psi_\lambda)\,dv=-\frac{D\lambda}{8\pi}\int_S\left(\psi_\lambda\frac{\partial\psi_n}{\partial n_s}-\psi_n\frac{\partial\psi_\lambda}{\partial n_s}\right)dS. \quad . \quad (9)$$

Here n_s denotes the outward normal at the particle surface into the region of integration, which is the inward normal to the surface.

Suppose that the dielectric constant of the colloidal particle is D^I and ψ^I represents the potential function inside the particle*. Since there is no charge inside the particle

$$\nabla^2\psi^I=0, \qquad \nabla^2\psi_n^I=0, \qquad \nabla^2\psi_\lambda^I=0, \quad . \quad . \quad . \quad . \quad (10)$$

and on the surface we have the boundary conditions

$$\psi=\psi^I, \qquad D\frac{\partial\psi}{\partial n_s}-D^I\frac{\partial\psi^I}{\partial n_s}=-4\pi\nu\lambda e_1. \quad . \quad . \quad . \quad . \quad (11)$$

We proceed to evaluate the various contributions to (6), which come from the surface integrals in (9). Interchanging the order of integration with respect to λ and S and then integrating by parts,

$$-\frac{D}{4\pi}\int_0^{\lambda'}d\lambda\int_S\psi_\lambda\frac{\partial\psi_n}{\partial n_s}dS=-\frac{D}{4\pi}\int_S\left[\psi\frac{\partial\psi_n}{\partial n_s}\right]_{\lambda=\lambda'}dS+\frac{D}{4\pi}\int_0^{\lambda'}d\lambda\int_S\psi\frac{\partial\psi_{n\lambda}}{\partial n_s}dS$$
$$. \quad . \quad . \quad (12)$$

It is seen from (11) that $\partial\psi/\partial n_s=0$ when $\lambda=0$, although it is possible that $\psi\neq0$. On applying the boundary conditions (11), and noting that $\nu_\lambda=0$, the right-hand side of (12) can be expressed as

$$\lambda' e_1\int_S\psi_{\lambda=\lambda'}\nu_n\,dS-\frac{D^I}{4\pi}\int_S\left[\psi^I\frac{\partial\psi_n^I}{\partial n_s}\right]_{\lambda=\lambda'}dS-e_1\int_0^{\lambda'}d\lambda\int_S\psi\nu_n\,dS$$
$$+\frac{D^I}{4\pi}\int_0^{\lambda'}d\lambda\int_S\psi^I\frac{\partial\psi_{n\lambda}^I}{\partial n_s}dS, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12i)$$

where $\psi_{\lambda=\lambda'}$ is the (surface) potential at stage λ' and position **s**. This disposes of the first term in the surface integral in (9). The second term in this surface integral yields the contribution

$$+\frac{D}{4\pi}\int_0^{\lambda'}\partial\lambda\int_S\psi_n\frac{\partial}{\partial\lambda}\left(\frac{\partial\psi}{\partial n_s}\right)dS=-e_1\int_0^{\lambda'}d\lambda\int_S\psi_n\nu\,dS+\frac{D^I}{4\pi}\int_0^{\lambda'}d\lambda\int_S\psi_n^I\frac{\partial\psi_\lambda^I}{\partial n_s}dS$$
$$. \quad . \quad . \quad (13)$$

to (*b*). If we now combine the three integrals in (12*i*) and (13) which involve the function ψ^I and make use of the first of the boundary conditions (11), then a simple integration by parts yields the expression

$$-\frac{D^I}{4\pi}\int_0^{\lambda'}d\lambda\int_S\left(\psi_\lambda^I\frac{\partial\psi_n^I}{\partial n_s}-\psi_n^I\frac{\partial\psi_\lambda^I}{\partial n_s}\right)dS=0. \quad . \quad . \quad . \quad . \quad (14)$$

* In the case of a metallic particle, we put ψ^I=constant.

This follows on application of Green's theorem and of the equation (10). Also, we have two terms, one from (12*i*) and the other from (13), which cancel out with the contribution from the surface integral on the right of (7). We finally derive that

$$\frac{\partial \mathrm{F}_e(\lambda', n)}{\partial n} = \lambda' e_1 \int_{\mathrm{S}} \psi_{\lambda=\lambda'} \nu_n \, d\mathrm{S}. \quad . \quad . \quad . \quad . \quad . \quad (15)$$

The formula (15) can be given a simple interpretation. When we bring up δn ions, each of charge $\lambda' e_1$, to the surface, the increment in the density of ions at position $\mathbf{s}$ is $\delta\nu = \nu_n \delta n = \nu_0(\lambda', \mathbf{s})\, \delta n$. The contribution to the change in electrical energy from the surface element $d\mathrm{S}$ is $\lambda' e_1 \psi_{\lambda=\lambda'} \delta\nu \, d\mathrm{S}$ which is the electrical work done in bringing $\delta\nu$ ions to the element $d\mathrm{S}$. In particular, if the density of ions is constant over the surface, then $\nu_0(\lambda', \mathbf{s}) = 1/\mathrm{S}$ and (15) becomes

$$\frac{\partial \mathrm{F}_e(\lambda, n)}{\partial n} = \lambda' e_1 \psi_{\mathrm{ave}}(\lambda', n), \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where $\psi_{\mathrm{ave}}(\lambda', n)$ is the average potential on the surface. If we define the average potential in terms of the weight function $\nu_n = \nu_0(\lambda', \mathbf{s})$, then (16) can be interpreted as covering the general case (15). We now integrate with respect to n from 0 to n', say, and so obtain

$$\mathrm{F}_e(\lambda', n') = \lambda' e_1 \int_0^{n'} \psi_{\mathrm{ave}}(\lambda', n) \, dn, \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where, in general,

$$\psi_{\mathrm{ave}}(\lambda', n) = \int_{\mathrm{S}} \psi_{\lambda=\lambda'} \nu_0(\lambda', \mathbf{s}) \, d\mathrm{S}. \quad . \quad . \quad . \quad . \quad . \quad (18)$$

The relation (17), which was already implied in a paper by Derjaguin (1940), was proposed without proof by the author in an earlier paper (Levine 1948). The right-hand side of (17) is the usual formula from electrostatics for the work done in bringing n' ions, each of charge $\lambda' e_1$, by infinitesimal amounts from the solution to the surface of the particle, the potential of which is $\psi_{\mathrm{ave}}(\lambda', n)$.

§ 3. Energy of two Particles.

We can readily extend the result (17) to a dilute sol containing P particles. Suppose that there are n'_p ions of type 1 on the surface of the pth particle. We imagine that these ions are transported from the solution to the surface at the same rate for all the particles and that the fraction ξ, which replaces the variable n, denotes the stage reached in this process. Thus $n_p = \xi n'_p$ will be the number of ions on particle p at stage ξ. The potential ψ_p and the surface density of ions ν_p on particle p will be considered as functions of λ, ξ and the position on the particle surface. The derivation of a simplified expression for the free energy follows the same steps as for a single particle provided we replace e_1 by $n'_p e_1$ for the pth

particle, differentiate with respect to ξ instead of n and sum over all the particles. The electrical free energy of the P double layers at stage λ' can be finally expressed in the form

$$\lambda' e_1 \sum_{p=1}^{P} n_p \int_0^1 \psi_p^{\text{ave}}(\lambda', \xi n_p')\, d\xi = \lambda' e_1 \sum_{p=1}^{P} \int_0^{n_p'} \psi_p^{\text{ave}}(\lambda', n_p)\, dn_p, \quad . \quad (19)$$

where $\psi_p^{\text{ave}}(\lambda', n_p)$ is the average potential on particle p.

In particular, we can apply this formula to two parallel plates, each of area A, at separation R. Since we are dealing with closed surfaces it is necessary to consider both sides of each plate and also the edges. The latter, however, can be neglected for sufficiently thin plates. In addition, we need not concern ourselves with the contribution to the free energy from the outer sides, since this will not vary with the separation R. Making use of equation (9) in a previous paper (Levine 1948) the electrical free energy due to the excess charge in the two inner double layers can be written as

$$\mathrm{F}_e(\lambda', \nu', \mathrm{R}) = 2\mathrm{A}\lambda' e_1 \int_0^{\nu'} \Phi(\lambda', \nu, \mathrm{R})\, d\nu, \quad . \quad . \quad . \quad . \quad (20)$$

where $\lambda' e_1 \Phi(\lambda', \nu, \mathrm{R})$ is the potential of the mean force acting on an adsorbed ion, due to the excess ions in the layers; the zero of this potential is taken in the interior of the solution. For finite plates, both ν and $\Phi(\lambda', \nu, \mathrm{R})$ would not be constant over the inner faces because of the edge effects and we would have to introduce averages. However, this correction is neglected here. The result (20) follows from the general theory based on the methods of statistical mechanics, which is developed elsewhere (Levine 1949b). If we now introduce the approximation of the Poisson-Boltzmann equation and identify $\Phi(\lambda', \nu, \mathrm{R})$ with the potential function $\psi_p(\lambda', n_p)$, $(p=1, 2)$, then the formula (20) is immediately derived from (19). In particular, if ν is sufficiently small, $\Phi(\lambda', \nu, \mathrm{R})$ can be assumed proportional to ν, in which case we have

$$\mathrm{F}_e(\lambda', \nu', \mathrm{R}) = \mathrm{A}\lambda' e_1 \nu' \Phi(\lambda', \nu', \mathrm{R}). \quad . \quad . \quad . \quad . \quad (20i)$$

This result was already suggested in an earlier paper (Levine 1948).

When the linear Debye-Huckel equation* is applied to two (infinite) parallel plates, the potential between the plates is given by

$$\psi(\lambda, x) = \frac{4\pi\nu e_1}{\mathrm{D}\kappa_0} \frac{\cosh \lambda\kappa_0 x}{\sinh \lambda\kappa_0 \mathrm{R}/2}, \qquad \kappa_0^2 = \frac{4\pi}{\mathrm{D}k\mathrm{T}} \sum_{i=1}^{s} n_i^0 e_i^2, \quad . \quad . \quad (21)$$

where x is measured from the median plane, so that we assume $\psi(\lambda, \mathrm{R}/2) = \Phi(\lambda, \nu, \mathrm{R})$. We find that an additional term appears on the right-hand side of $(20i)$, namely $-8\pi\mathrm{A}\nu'^2 e_1^2/\mathrm{D}\kappa_0^2\mathrm{R}$. This difficulty was met in a previous paper (Levine 1946) and a lengthy discussion by Overbeek (1946) and the author (Levine 1946a) followed concerning this

* We expand $f(\lambda\psi)$ in powers of ψ and retain the first term only, namely $f(\lambda\psi) \approx \lambda\kappa_0^2\psi$.

discrepancy. Although the essential basis for the appearance of this extraneous term was first proposed by Overbeek, no conclusive proof that (20*i*) is actually the correct result was given. We are now in a position to show clearly the origin of this additional term.

When all the ions are discharged, the formula (21) yields the condition

$$\lim_{\lambda=0} \lambda\psi(\lambda, x)=8\pi\nu e_1/\mathrm{D}\kappa_0^2\mathrm{R}=m, \text{ say,} \quad . \quad . \quad . \quad . \quad (22)$$

throughout the region between the two plates and the fact that the potential $\psi(\lambda, x)\to\infty$ as $1/\lambda$ when $\lambda\to 0$ creates a difficulty. Use of the Boltzmann distribution law yields the expression $n_i^0 \exp(-e_i m/k\mathrm{T})$ for the limiting density (at $\lambda=0$) of i ions at position x. But if the short-range van der Waals forces near the particle walls are neglected, then necessarily the completed discharged ions should be uniformly distributed throughout the dispersion medium, which means that the condition $\lim_{\lambda=0} \lambda\psi(\lambda, x)=0$ must be satisfied, in contradiction to (22).

The analysis for a single plate developed in the previous section can be applied to each plate, provided the volume of the diffuse outer layer does not extend to infinity but to the median plane. If now $\lim_{\lambda\to 0} \lambda\psi\neq 0$ then the integral in (5) still converges but this is not true of each of the separate terms obtained by substituting (4) into (5). To overcome this difficulty of divergent integrals, we replace the lower limit of integration $\lambda=0$ by $\lambda=\epsilon$, where ϵ is a small positive number, and after integration let $\epsilon\to 0$. An additional term is obtained on the right of (12), namely*

$$-\frac{\mathrm{D}}{4\pi}\lim_{\lambda\to 0}\int_s \psi\frac{\partial\psi_n}{\partial n_s}\,d\mathrm{S}=-\lim_{\lambda\to 0}\lambda e_1\psi_{\mathrm{ave}}(\lambda, n), \quad . \quad . \quad . \quad (23)$$

which does not vanish, since $\partial\psi/\partial n_s$ behaves as λ whereas ψ behaves as $1/\lambda$ when $\lambda\to 0$. The form on the right of (23) is obtained when we apply the condition that $\mathrm{F}_e(\lambda, n)\to 0$ as $\lambda\to 0$. We now need to introduce a corresponding correction to the right-hand side of our basic relation (17), which becomes

$$\mathrm{F}_e(\lambda', n')=\lambda' e_1\int_0^{n'}\psi_{\mathrm{ave}}(\lambda', n)\,dn-\lim_{\lambda\to 0}\lambda e_1\int_0^{n'}\psi_{\mathrm{ave}}(\lambda, n)\,dn. \quad . \quad (24)$$

The second term on the right of (24) was suggested by Overbeek (1946) although his explanation was not quite adequate. In the case of each of our two parallel plates, this new term becomes

$$-\frac{\mathrm{DA}}{4\pi}\lim_{\lambda\to 0}\int_0^{\nu'}\psi\left(\lambda, \frac{\mathrm{R}}{2}\right)\left(\frac{\partial\psi_\nu(\lambda, x)}{\partial x}\right)_{x=\mathrm{R}/2}d\nu$$

$$=-\lim_{\lambda\to 0}\mathrm{A}\lambda e_1\int_0^{\nu'}\psi\left(\lambda, \frac{\mathrm{R}}{2}\right)d\nu=-\frac{4\pi\mathrm{A}\nu'^2 e_1^2}{\mathrm{D}\kappa_0^2\mathrm{R}}, \quad . \quad . \quad . \quad (25)$$

* We note that at the median plane, $\partial\psi/\partial n_s=0$, since the potential is a minimum.

which is precisely half the additional (extraneous) term appearing on the right of (20*i*). The anomaly arises because we are considering two plates of infinite extent. This model cannot be used to describe two finite plates in the neighbourhood of $\lambda=0$ since it prevents the diffusion of the excess charge of the diffuse layers between the plates into the bulk of the solution as $\lambda\to 0$. The necessary condition $\lim_{\lambda\to 0}\lambda\psi=0$ ensures that the expression (25) always vanishes and hence our formula (17) applies to each of the finite plates.

For two identical particles, symmetrically orientated with respect to the line joining their centres of mass, we shall denote the (weighted) average potential over each particle, which is defined in (18), by $\psi_{\text{ave}}(\lambda', n, \text{R})$. It follows from (17) that the electrical free energy of the double layers of the two particles is given by

$$\text{F}_e(\lambda', n', \text{R})=2\lambda' e_1 \int_0^{n'} \psi_{\text{ave}}(\lambda', n, \text{R})\, dn. \quad . \quad . \quad . \quad . \quad (26)$$

This is the simplified form that we set out to derive in this section.

§4. Casimir's Theorem of Equivalence.

This theorem of the equivalence of the two methods, employed by Verwey and Overbeek (1948) to evaluate the free energy of the double layer, can be derived from our general results in §2. It will be sufficient to consider a single particle, since the argument is readily extended to an arbitrary number of particles. Also we shall first derive the Casimir theorem in the special circumstances when (16) is replaced by

$$\frac{\partial \text{F}_e(\lambda', n)}{\partial n}=\lambda' e_1 \psi(\lambda', n), \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where $\psi(\lambda', n)$ denotes the constant potential at the surface, *i. e.* it does not depend on the position **s,** but it is a function of λ' and n. It should be noted that n is treated as an arbitrary parameter in (27). The following considerations show that (27) is valid for any n in two particular cases.

Suppose that we replace the general form (1) for ν by

$$\nu=n^*(\lambda')\nu_0^*(\lambda', \mathbf{s}), \qquad \int_{\text{S}} \nu_0^*(\lambda', \mathbf{s})\, d\text{S}=1, \quad . \quad . \quad . \quad . \quad (1i)$$

such that in (18), $\psi_{\lambda=\lambda'}$ is uniform over the particle surface. In these circumstances, it might appear that (27) applies, but we find that in general, this is not the case. For let the arbitrary function $\nu_0(\lambda', \mathbf{s})$ in (18) be replaced by $\nu_0^*(\lambda', \mathbf{s})$ but at the same time n be given any value. Then the surface potential $\psi_{\lambda=\lambda'}$ will depend on n and, in general, also on the position **s**. If we now introduce $n=n^*(\lambda')$ as well as the distribution $\nu_0^*(\lambda', \mathbf{s})$, then $\psi_{\lambda=\lambda'}$ does become uniform over the surface. It follows that for this particular value of n (and distribution $\nu_0^*(\lambda', \mathbf{s})$), $\psi_{\text{ave}}(\lambda', n^*(\lambda'))$ can be identified with $\psi(\lambda', n^*(\lambda'))$; in other words the relation (27) becomes valid if we susbstitute $n=n^*(\lambda')$. The problem is whether $\psi_{\lambda=\lambda'}$ is

constant over the surface when the relative distribution is again $\nu_0^*(\lambda', \mathbf{s})$ but n is arbitrary. One case where this is true is that of small n, when (2) reduces to the linear Debye-Huckel equation, with solution $\psi = n\phi$, where the function ϕ depends on the form of $\nu_0^*(\lambda', \mathbf{s})$ but not on n. For we can choose the distribution $\nu_0^*(\lambda', \mathbf{s})$ such that the value of ϕ at the surface (at stage λ') is constant, independent of the position $\mathbf{s}$. Denoting this constant value by $\phi_{\lambda=\lambda'}$, (27) will apply, with $\psi(\lambda', n) = n\phi_{\lambda=\lambda'}$. However, if n is not small, the essential property that ψ be proportional to n no longer holds and (27) is not valid for a particle of general shape. The other case arises when we are dealing with a spherical, cylindrical or plate-like particle, or with two parallel plates, provided the surface density of ions is uniform (equal to n/S) and any edge effects due to the finite size of the particles are neglected.

Now in §2, we can treat λ' as a parameter, just like n, and we obtain the relations

$$\nabla^2\psi_{\lambda'} = \lambda^2 f'\psi_{\lambda'}, \qquad \nabla^2\psi^{\mathrm{I}}_{\lambda'} = 0,$$

where the subscript λ' denotes partial differentiation. Using precisely the same method as in §2, we can derive quite generally

$$2\int_0^{\lambda'} \frac{1}{\lambda}\frac{\partial \mathrm{E}(\lambda, n\,;\,\lambda')}{\partial\lambda'}\,d\lambda = \lambda' e_1 \int_{\mathrm{S}} \psi_{\lambda=\lambda'}\nu_{\lambda'}\,d\mathrm{S} = \lambda' e_1 \Psi(\lambda', n), \text{ say.} \qquad (28)$$

It is observed that (28) immediately follows from (6) and (15) when n is replaced by λ'. When (27) applies, then $\psi(\lambda', n) = \psi_{\lambda=\lambda'}$ and hence

$$\Psi(\lambda', n) = \psi(\lambda', n)\frac{\partial}{\partial\lambda'}\left[\int_{\mathrm{S}} \nu\,d\mathrm{S}\right] = 0, \quad . \quad . \quad . \quad . \quad (28i)$$

since the surface integral in (28i) is equal to n.

Since it is simpler to treat the case where the potential at the surface is constant (equal to $\psi(\lambda', n)$), we begin with the relation

$$2\int_0^{\lambda} \frac{1}{\lambda}\mathrm{E}(\lambda, n'\,;\,\lambda')\,d\lambda = \lambda' e_1 \int_0^{n'} \psi(\lambda', n)\,dn, \quad . \quad . \quad . \quad (29)$$

which follows from (5) and (27). Remembering that λ' and n' are considered to be independent of one another, we differentiate both sides of (29) partially with respect to λ', make use of the relations (28) and (28i), and then put $\lambda' = \lambda$. In this way we obtain

$$\frac{2}{\lambda}\mathrm{E}(\lambda, n'\,;\,\lambda) = e_1\int_0^{n'} \psi(\lambda, n)\,dn + \lambda e_1 \int_0^{n'} \psi_\lambda(\lambda, n)\,dn. \quad . \quad . \quad (30)$$

This is an identity in λ and n', and in particular, will be true if n' is chosen that function of λ which is defined by requiring that $\psi(\lambda, n') = \psi_0$, a constant. This function is given by the inverse relation which we write as

$$n' = n(\lambda, \psi_0) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

If the number of surface ions is given by (31), the electrostatic energy of the double layer can be expressed as

$$\mathrm{E}(\lambda, n(\lambda, \psi_0)\,;\ \lambda)=\tfrac{1}{2}\lambda e_1\psi_0 n(\lambda, \psi_0)+\tfrac{1}{2}\int_{\mathrm{V}}[\psi\rho]_{\psi_0}\,dv, \quad . \ . \quad (32)$$

where the surface potential is constant (equal to ψ_0) when integrating over the volume V on the right-hand side. If we also substitute (31) for n' into (30), then it follows that

$$\frac{1}{\lambda}\int_{\mathrm{V}}[\psi\rho]_{\psi_0}dv=-e_1\int_0^{\psi_0} n(\lambda, \psi)\,d\psi+\lambda e_1\int_0^{n(\lambda, \psi_0)}\psi_\lambda(\lambda, n)\,dn. \quad . \quad (33)$$

The first term on the right of (33) has been written in place of

$$e_1\int_0^{n(\lambda, \psi_0)}\psi(\lambda, n)\,dn-e_1\psi_0 n(\lambda, \psi_0).$$

Since $\psi_0=\psi(\lambda, n(\lambda, \psi_0))$, we have the relations

$$\int_0^{n(\lambda, \psi_0)}\psi_\lambda(\lambda, n)\,dn=\frac{\partial}{\partial\lambda}\left[\int_0^{n(\lambda, \psi_0)}\psi(\lambda, n)\,dn-\psi_0 n(\lambda, \psi_0)\right]=-\frac{\partial}{\partial\lambda}\int_0^{\psi_0} n(\lambda, \psi)\,d\psi,$$

and hence (33) becomes

$$\frac{1}{\lambda}\int_{\mathrm{V}}[\psi\rho]_{\psi_0}\,dv=-\frac{\partial}{\partial\lambda}\left[\lambda e_1\int_0^{\psi_0} n(\lambda, \psi)\,d\psi\right]. \quad . \ . \ . \quad (34)$$

Integrating with respect to λ from 0 to λ', we finally derive

$$\int_0^{\lambda'}\frac{d\lambda}{\lambda}\int_{\mathrm{V}}[\psi\rho]_{\psi_0}\,dv=-\lambda' e_1\int_0^{\psi_0} n(\lambda', \psi)\,d\psi+\lim_{\lambda\to 0}\lambda e_1\int_0^{\psi_0} n(\lambda, \psi)\,d\psi. \quad (35)$$

For plate-like particles, the second term on the right of (35) vanishes and we obtain the formula derived by Casimir (1948). For a cylindrical particle, the solution of the linear Debye-Huckel equation yields the formulæ

$$\psi(\lambda, n)=\frac{2ne_1}{\mathrm{D}\kappa_0 a}\frac{\mathrm{K}_0(\lambda\kappa_0 a)}{\mathrm{K}_1(\lambda\kappa_0 a)}, \qquad n(\lambda, \psi_0)=\frac{\mathrm{D}\kappa_0 a\psi_0}{2e_1}\frac{\mathrm{K}_1(\lambda\kappa_0 a)}{\mathrm{K}_0(\lambda\kappa_0 a)}, \quad . \quad (36)$$

where n is the number of surface ions per unit length of particle, a is the radius of the circular cross-section and K_0 and K_1 are Bessel functions of purely imaginary argument. Since $\mathrm{K}_0(\lambda\kappa_0 a)$ and $\mathrm{K}_1(\lambda\kappa_0 a)$ behave as $\ln\lambda$ and $1/\lambda\kappa_0 a$ respectively when $\lambda\to 0$, the second term on the right of (35) again vanishes, although $n\to\infty$ (which is, of course, physically impossible).

In the case of a spherical particle of radius a, use of the linear Debye-Huckel equation for ψ yields the formula

$$\psi=\psi_0 a\exp\left[-\lambda\kappa_0(r-a)\right]/r, \quad . \ . \ . \ . \ . \ . \ . \quad (37)$$

for the potential distribution in the double layer, where r is the distance from the centre of the sphere. Thus

$$\psi(\lambda, n)=\frac{n\lambda e}{\mathrm{D}a(1+\lambda\kappa_0 a)}, \qquad n(\lambda, \psi_0)=\frac{\mathrm{D}a\psi_0}{\lambda e_1}(1+\lambda\kappa_0 a), \quad . \ . \quad (38)$$

so that $n(\lambda, \psi_0) \to \infty$ as $\lambda \to 0$. We now find that

$$\lim_{\lambda \to 0} \lambda e_1 \int_0^{\psi_0} n(\lambda, \psi)\, d\psi = D a \psi_0^2/2. \quad \ldots\ldots (39)$$

If we substitute (37) and the corresponding volume density of charge

$$\rho = -(D/4\pi)\nabla^2\psi = -D\lambda^2\kappa_0^2\psi/4\pi, \quad \ldots\ldots (40)$$

which follows from (2) and the Debye-Huckel approximation $f(\lambda\psi) = \lambda\kappa_0^2\psi$, the left-hand side of (35) becomes

$$-D a^2\psi_0^2 \int_0^{\lambda'} \lambda \exp(2\lambda\kappa_0 a)\, d\lambda \int_a^{\infty} \exp(-2\lambda\kappa_0 r)\, dr = -(D a\psi_0^2/2)\lambda'\kappa_0 a. \quad (41)$$

However, it follows from (38) that the first term on the right of (35) is

$$-(D a\psi_0^2/2)(1+\lambda'\kappa_0 a). \quad \ldots\ldots (42)$$

No mention of this exception to the Casimir theorem is made in the book of Verwey and Overbeek. The omission by Casimir (1948) of the second term on the right of (35) is made at the end of his proof when he assumes (in his notation) $\partial\chi(\lambda=0)/\partial n = 0$, *i. e.* the derivative of the potential in the direction normal to the particle surface vanishes at $\lambda = 0$ *. However, this condition is not satisfied for a spherical particle since it follows immediately from (37) that $\lim_{\lambda \to 0} (\partial\psi/\partial r)_{r=a} = -\psi_0/a$. This derivative of the potential does vanish at the surface of a plate or cylinder.

It is possible to obtain a generalization of the relation (35) for a particle of any shape, the surface of which need not be at a constant potential. On the right-hand side of (29) we replace the function $\psi(\lambda', n)$ by $\psi_{\text{ave}}(\lambda', n)$ which is defined by (18), with the distribution $\nu_0(\lambda', \mathbf{s})$ chosen arbitrarily. Differentiating the modified form of (29) with respect to λ', we again put $\lambda' = \lambda$. It is now assumed that n' is chosen, such a function of λ that the quantity $\psi_{\text{ave}}(\lambda, n')$ becomes equal to a constant ψ_0, independent of λ. This function will be denoted as

$$n' = n_{\text{ave}}(\lambda, \psi_0), \quad \ldots\ldots (31i)$$

which replaces (31). (Since $\psi_{\text{ave}}(\lambda, n')$ refers to stage λ, it is defined by the weight function $\nu_0(\lambda, \mathbf{s})$ in (18).) The operations performed on the right-hand side of (29) are precisely the same as those in the analysis above. Turning our attention to the left-hand side of (29), it is seen that we are charging the ions from 0 to λ', subject to a fixed surface distribution of ions, given by $\nu = n'\nu_0(\lambda', \mathbf{s})$. Also, an additional term now appears on the left of (30), namely $\lambda e_1 \Psi(\lambda, n')$, which will not vanish in the general case. When we reach the equation which is the generalization of (35), namely

$$\int_0^{\lambda'} \frac{d\lambda}{\lambda} \int_V [\psi\rho]_{\psi_{\text{ave}} = \psi_0} dv = -e_1 \int_0^{\lambda'} \Psi(\lambda, n_{\text{ave}}(\lambda, \psi_0))\, d\lambda - \lambda' e_1 \int_0^{\psi_0} n_{\text{ave}}(\lambda, \psi)\, d\psi$$

$$+ \lim_{\lambda \to 0} \lambda e_1 \int_0^{\lambda_0} n_{\text{ave}}(\lambda, \psi)\, d\psi, \quad \ldots\ldots (43)$$

* Subject to fixed surface potential (4_0).

the integration with respect to λ on the left-hand side is being carried out under the condition that, at each stage λ, the average potential over the particle surface is $\psi_{ave}(\lambda, n_{ave}(\lambda, \psi_0))=\psi_0$, a constant. This result reduces to the Casimir theorem for a particle of general shape in the following special circumstances. The distribution $\nu_0(\lambda, \mathbf{s})$ (at stage λ) and the total number of surface ions n are chosen in such a way that the surface potential is constant over the surface and equal to ψ_0. We can denote this distribution by ν_0^* $(\lambda, \mathbf{s})$ and the value of n by $n^*(\lambda, \psi_0)$, which now replaces $n_{ave}(\lambda, \psi_0)$. Then it follows from (28) that $\Psi(\lambda, n^*(\lambda, \psi_0))=0$ and the relation (18) yields (at $\lambda'=\lambda$) $\psi_{ave}(\lambda, n^*(\lambda, \psi_0))=\psi_{\lambda=\lambda}=\psi_0$. Hence (43) takes on the same form as (35), which is the required result.

§ 5. Discussion.

We can suggest the following physical interpretation of our theorem which is expressed in equations (5), (17) and (18). All methods of picturing the formation of the electric double layers of the particles should lead to the same value for the free energy provided such methods describe thermodynamically reversible processes. Suppose now that the surface density of ions is treated as a thermodynamic variable, which we can change at will. Then the charging process introduced in the definition (5) of the free energy is carried out under equilibrium (reversible) conditions and is therefore valid. Our second method which leads to the expression (17) for the free energy is also valid provided the electric work done in bringing the ions to the surface is against the potential of the mean force acting on the ions. When we use the approximation of the Poisson-Boltzmann equation (the potential of the mean force being identified with the mean potential) and also restrict ourselves to very dilute sols, then this additional condition in the second method is apparently accounted for.

It might appear that we have introduced some restriction by keeping the distribution (and hence number) of surface ions fixed when charging the double layers. However, this is not the case and the explanation depends on two theorems in (classical) statistical mechanics. Firstly, it has been shown by Onsager (1933) that it is permissible to apply the charging process to a sub-space of the positional part of the phase space of the ions. Such a sub-space is defined by a fixed (but arbitrary) distribution of surface ions but *not*, in general, by choosing the surface potential as constant. Secondly, the free energy has an extremely sharp minimum at the equilibrium distribution of surface ions. This latter property implies that the size of the particles is so large that we are dealing with a macrospocic system. By treating a large number of particles we have, in effect, macroscopic particle surfaces. (We ignore here the complication which arises when the adsorbed ions form part of the crystal lattice of the particles and which is briefly discussed elsewhere (Levine 1949a)).

The two methods of Verwey and Overbeek for evaluating the free energy lead to equation (35), with the second term on the right-hand side absent. In developing a physical argument to interpret the equivalence of their methods, these authors introduce the apparently arbitrary assumption that the charging process be carried out under the condition of constant surface potential. Previously (Levine 1946b) the author had raised the question as to whether the equilibrium conditions were being maintained in this charging method. In the case of spherical particles, for example, the number of surface ions n becomes infinite as $\lambda \to 0$, when the surface potential is kept fixed. As a consequence, the two methods originally proposed by Verwey and Overbeek yield different results and the Casimir theorem of equivalence has to be modified by introducing the second term on the right of (35).

It appears, therefore, that equilibrium conditions are not kept in the process of charging at constant surface potential. In the opinion of the author, little or no physical significnace can be attached to the artificial process visualized by Verwey and Overbeek. Nevertheless it is possible to obtain the correct expression for the free energy by this method, although not in all cases (*e. g.* spheres). The reason for the success of this approach is to be found in the nature of the theorems relating to the free energy, which are based on the use of the Poisson-Boltzmann equation and the restriction to very dilute sols. A number of these theorems have been developed in this paper, and others will be given later. Thus, we regard the equivalence theorem of Casimir as a convenient mathematical relation.

In conclusion, the author wishes to thank Professor J. Th. G. Overbeek for helpful correspondence on a number of aspects in this paper.

References.

Casimir, H. G. B., 1948, in *Theory of Stability of Lyophobic Colloids, Elsevier* p. 63, by Verwey and Overbeek.

Derjaguin, B., 1940, *Trans. Faraday Soc.*, **36**, 203.

Grimley, T. B., and Mott, N. F., 1947, *Faraday Soc. Discussion*, **1**, 3.

Levine, S., 1946a. *Trans. Faraday Soc., Swelling and Shrinking issue*, **42B**, 102; 1946b, *Ibid.*, 128; 1948, *Trans. Faraday Soc.*, **44**, 833; 1949a, *J. Colloid Science* (in press); 1949b, submitted to *Proc. Camb. Phil. Soc.*

Onsager, L., 1933, *Chem. Rev.*, **13**, 73.

Overbeek, J. Th. G., 1946, *Trans. Faraday Soc., Swelling and Shrinking issue*, **42B**, 126.

Verwey, E. J. W. and Overbeek, J. Th. G., 1948, *Theory of the Stability of Lyophobic Colloids, Elsevier.*

V. *Noise in Ionization Chamber Pulse Amplifiers.*

By R. Wilson, D.Phil.,
Clarendon Laboratory, Oxford*.

[Received September 15, 1949.]

Abstract.

The general problem of the detection of heavy ionizing particles, in the presence of γ radiation and background noise from the amplifier, is considered. It is shown that the maximum possible signal to noise ratio is obtained if the input pulse is shaped to a form $\exp\{-|t-T_D|/\tau\}$. The signal to noise ratio is also calculated for other common pulse shapes.

A practical amplifier circuit is discussed which may be used to detect 100 kv. energy protons in the presence of intense γ radiation.

Introduction.

The signal from an ionization chamber is a pulse of current which flows in an external circuit as the movement of ions induces charge on the electrodes. If free electrons are collected, and the counter geometry is properly designed, this time may be as low as $0{\cdot}1\mu$s. which is less than the usual amplifier time constants. The total charge which passes is proportional to the energy of the initial ionizing particle.

It is desired to amplify this current pulse and to distinguish it from background " noise ". This background noise is due to the superposition of many smaller pulses ; some due to electrons in the counter, others due to the spontaneous thermal agitation of electrons in the electrical circuit (Johnson noise) or the spontaneous movement of electrons in the amplifying valves (shot effect).

The wanted current pulse, like all other current pulses, is transformed into a step pulse of potential on the grid of the first valve, by charging the stray capacity to earth of the ionization chamber and valve. The grid leak of this valve discharges the condenser after a long time, but the grid leak is usually made large in order not to produce appreciable resistor noise, so that this time is sufficiently long to be regarded as infinite.

The amplifier will distort this pulse of potential. Part of this distortion may be deliberate, part caused by defects in the amplifier. It is convenient, however, to consider an ideal, non-distorting, linear amplifier and certain linear pulse-shaping networks which distort this and every other pulse. Then the problem is to decide upon the best way to distort this pulse in order to discriminate against the background noise. It is convenient

---

* Communicated by the Author.

to " refer " the various sources of background noise to the input as an " equivalent noise charge ". This is a fictitious noise charge on the ionization chamber electrodes, which produces the same signal at the amplifier output as the noise.

It will be shown that with a conventional resistance—capacitance pulse shaping network the best signal to noise ratio is obtained when the time constants of " differentiation " and " integration " are equal ; that an improvement of 20 per cent is possible by using a shorted delay time for shaping, and that a further improvement of 15 per cent is possible by using a network to shape the pulse to a form $\exp\{-|t-T_D|/\tau\}$. The overall improvement of 35 per cent may be very important.

Analytical Expression for Signal to Noise Ratio.

It will here be assumed that the duration of the current pulse is short compared with the pulse shaping time constants so that the input may be regarded as a function $Q\delta(t)$ of time—where $\delta(t)$ is Dirac's δ function defined by $\delta(t)=0$ for $t\neq 0$ and $\int_{-\infty}^{+\infty}\delta(t)\,dt=1$.

If this assumption is not true (as for example with ion collection in counters) the pulses from the ionization chamber will be attenuated more than the background pulses which is undesirable if a high signal to noise ratio is wanted.

The potential charge on the first valve grid is then $\frac{Q}{c}H(t)$ where $H(t)$ is Heaviside's unit function.

$$H(t)=1 \qquad t>0,$$

$$H(t)=0 \qquad t<0,$$

$$\text{and } \frac{d\{H(t)\}}{dt}=\delta(t).$$

All pulses will be modified by the shaping network. If the input potential pulse is $H(t)$, then the output pulse is a function $f(t)$ of the shaping network. If the input potential pulse is $\delta(t)$, the output pulse will be $d/dt\{f(t)\}=f'(t)$.

The amplifier noise is due to the superposition of the effects of individual electrons in those amplifier elements which are at a low signal level—the input grid leak, the first valve and the first valve anode load. When the ionization chamber is used to count heavy ionizing particles in the presence of intense γ radiation the secondary electron pulses in the chamber also cause a background.

The noise is often expressed in terms of the amplifier bandwidth, here this practice is reversed, and the noise is expressed in terms of the output pulse shape, which is considered to be more fundamental in a pulse amplifier than the bandwidth. This involves a more fundamental consideration of the noise problem.

The valve shot noise may be directly expressed as the superposition of pulses caused by individual electrons in the current stream. Each electron will produce a current pulse as it reaches the electrode ; since the transit time ($\sim 10^{-9}$ secs.) is small compared with the amplifier time constants, this may be regarded as a pulse of $e\delta(t)$.

Here can be distinguished two classes of noise. The valve anode has a resistive load R, so that a potential pulse $Re\delta(t)$ is developed across the load. The superposition of these pulses will be called class I noise. On the other hand, the valve grid has a capacitative load C, and a potential pulse $e\mathrm{H}(t)/\mathrm{C}$ will be developed across the load. The superposition of these other pulses will be called class II. noise.

In practice it is desired to reduce the signal level at which accurate counting is possible. An arbitrary maximum rate of counting of noise peaks is set for each experiment, and the best amplifier conditions must be found in order to reduce the signal level at which this counting rate occurs. A useful concept for this problem is the root mean square noise fluctuation. If the noise current is $\mathrm{I}(t)$ the r.m.s. noise fluctuation σ is given by $\sqrt{\overline{(\mathrm{I}(t)-\overline{\mathrm{I}(t)})^2}}$.

If the rate of arrival of noise pulses is large compared with the frequencies passed by the amplifier, the rate of counting of noise peaks may be expressed in terms of σ, the signal level V, and the amplifier bandwidth. (See equation (20) later). The problem of minimizing the signal level for a constant rate of counting is difficult, but the amplifier bandwidth only enters into the counting rate in a linear term, whereas σ enters in an exponential term. A close approximation, increasingly accurate as V/σ increases, will be to minimize σ. A rough check shows that the pulses should be about 10 per cent wider than given by this criterion, but this would alter the minimum detectable signal by only 1 per cent.

The assumption that the rate of arrival of noise pulses will be greater than the frequencies passed by the amplifier will usually hold for shot noise from both the anode and the grid, and for thermal (resistor) noise. But the γ ray background may not be sufficiently intense for the pulses to " pile up " in this way. In this case the γ rays will be counted as separate pulses of height ke.

Change of pulse shape will not then alter the rate of counting of background at all ; so that the γ ray background must be treated as distinct from all other forms of noise. If the pulses are small, their effect will be inappreciable, and if large, they will be similar to unwanted heavy ionizing particles.

An intermediate case will often arise where a certain amount of pile-up occurs which is difficult to treat theoretically, especially since the other forms of noise are added.

If these limitations do not apply, we may proceed by finding the root mean square fluctuation from Campbell's Theorem (Campbell 1909).

Campbell's theorem states that if a noise current $\mathrm{I}(t)$ is caused by the

superposition of ν events per second, each causing an effect $f(t)$ then the root mean square fluctuation σ is given by

$$\sigma^2=\overline{(I(t)-\overline{I(t)})^2}$$

$$=\nu\int_0^\infty \{f(t)\}^2\,dt. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The anode current develops a potential difference across a resistance and the pulse $\delta(t)$ later is transformed into a pulse $f'(t)$ by the pulse shaping network. We therefore derive, following Schottky (1918),

$$\sigma_I^2=R^2e^2\nu\int_0^\infty \{f'(t)\}^2\,dt$$

$$=R^2I_ae\int_0^\infty \{f'(t)\}^2\,dt. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Where σ_I is the r.m.s. anode current fluctuation. If we refer this to the ionization chamber input and put in the usual modification for space change limitation (North 1940), we obtain

$$\sigma_Q^2=\frac{0{\cdot}12ec^2}{g_m}\frac{I_a}{I_a+I_{sg}}\left\{1+\frac{8I_{sg}}{g_m}\right\}\int_0^\infty \{f'(t)\}^2\,dt*$$

(where c is the total input capacity of valve and counter) . (3)

for conventional valves with oxide-coated cathodes and a high amplification factor.

The grid current produces a potential difference across a condenser and the fluctuation becomes

$$\sigma_Q^2=I_ge\int_0^\infty \{f(t)\}^2\,dt. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The background secondary electrons in the counter will produce a background

$$\sigma_Q^2=I_\gamma(ke)\int_0^\infty \{f(t)\}^2\,dt. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Where $I\gamma$ is the total ionization current due to γ rays and ke is the average charge collected per electron. The noise in the input grid leak is (following Johnson 1925)

$$\sigma_Q^2=2kT/R\int_0^\infty \{f(t)\}^2\,dt. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

which is small compared with the grid current noise if R is large, or more exactly if

$$2kT/R \ll (I_g+kI_\gamma)e.$$

It can be shown that the anode resistor noise belongs to the same "class" as the anode current noise and is always small in comparison.

---

* All r.m.s. fluctuations are now expressed in terms of a fluctuation of charge on the first valve grid. This is denoted by use of the suffix σ_Q.

Flicker noise cannot be classified in either class I. or class II., but is intermediate. The original process is frequency dependent and σ^2 falls inversely as the first power of the frequency. It will only be appreciable for pulse widths greater than 1 millisecond.

Induced grid noise is of the same class (class I.) as anode noise, but is always negligible in comparison. All the usual forms of noise can therefore be represented in the form

$$\sigma_Q^2 = A\int_0^\infty \{f(t)\}^2\,dt + B\int_0^\infty \{f'(t)\}^2\,dt \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where

$$A = (I_g + kI_\gamma)e + 2kT/R, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7a)$$

$$B = \frac{0{\cdot}12eC^2}{g_m}\,\frac{I_a}{I_a + I_{sg}}\,\{1 + 8I_{sg}/g_m\}. \quad . \quad . \quad . \quad . \quad . \quad (7b)$$

For a triode, B reduces to $\dfrac{0{\cdot}12eC^2}{g_m}$.

Optimum Pulse Shape.

It is often of interest to obtain the maximum possible signal to noise ratio with no restrictions on the shape of the output pulse. This may be simply derived from the expression above by minimizing σ_Q^2 for a constant peak value of $f(t)$. $f(t)$ must clearly be zero at $t=0$ and $t=\infty$ the first from initial conditions, and the second to make

$$\int_0^\infty \{f(t)\}^2\,dt$$

finite.

Moreover it should be a symmetrical pulse; the integral $\int_0^\infty \{f(t)\}^2\,dt$ can be separated into

$$\int_0^{T_D} \{f(t)\}^2\,dt + \int_{T_D}^\infty \{f(t)\}^2\,dt,$$

where $f(t)$ is a maximum at $t=T_D$. If $f(t)$ is symmetrical the two integrals will be equal; if $f(t)$ is not symmetrical and

$$\int_0^{T_D} \{f(t)\}^2\,dt > \int_{T_D}^\infty \{f(t)\}^2\,dt,$$

then σ will be made smaller by altering $f(t)$ between 0 and T to make the pulse symmetrical.

If $f(t)$ is a function of (t/τ) alone where τ is some variable, then the best choice of τ is given when the two integrals which add to produce σ^2 are equal, *i. e.* when

$$A\int_0^\infty \{f(t)\}^2\,dt = B\int_0^\infty \{f'(t)\}^2\,dt. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

If we can choose the form of the function $f(t)$ then the best function is given by

$$f(t)=\exp\{-|t-T_D|/\tau\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

and $\tau^2=B/A$.

This follows by minimizing

$$\int_{T_D}^{\infty} A\{f(t)\}^2\,dt+\int_{T_D}^{\infty} B\{f'(t)\}^2\,dt=\sigma^2/2$$

with the auxiliary conditions, that at

$$t=\infty \quad f(t)=0,$$

$$t=T_D \quad f(t)=1,$$

this is normalizing the noise to unit pulse height. Then using the Calculus of Variations we obtain

$$2f(t)=\frac{B}{A}\frac{d}{dt}\{2f'(t)\} \quad \text{(Euler's formula)},$$

$$f(t)=\frac{B}{A}f''(t),$$

whence $f(t)=\exp\{-(t-T_D)/\tau\}$ where for $t>T_D$ and $f(t)=\exp\{(t-T_D)/\tau\}$ for $t<T_D$ by the symmetry property discussed above. There is a discontinuity in $f'(t)$ at $t=T_D$ but no discontintinuity in $\{f'(t)\}^2$.

Clearly

$$A\int_0^{\infty}\{f(t)\}^2\,dt=B\int_0^{\infty}\{f'(t)\}^2\,dt$$

in agreement with formula (8).

From a physical point of view, it is of interest to see why this pulse shape is the best. The sharp pointed peak will not give use to excessive noise, since the extra height adds little to the area under the curve of $\{f(t)\}^2$, whereas it adds considerably to the signal. This pointed peak is the great advantage.

Comparison of Pulse Shapes.

It remains to be seen how much improvement can be obtained in practice by using this pulse shape in preference to other pulse shapes; in order to do this it is necessary to consider the response of commonly used amplifier networks to a pulse $H(t)$ or $\delta(t)$ to obtain the shape $f(t)$ and $f'(t)$.

The response of the amplifier to a function $H(t)$ or $\delta(t)$ may be readily found by the method of the Laplace Transformation.

$$F(p)=\int_0^{\infty} e^{-pt}f(t)\,dt. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

The transform of $\delta(t)$ is 1 and of $H(t)$ is $1/p$.

The system function for the amplifier $G(p)$ is written down by replacing the differential operator d/dt by p. Then the response to a pulse $\delta(t)$ is the inverse transform of $G(p)$ and the response to a pulse $H(t)$ is the inverse transform of $G(p)/p$. The method is discussed by Elmore (1948) and Carslaw and Jaeger (1941). The method is here applied to obtain the signal to noise ratio for a variety of pulse-shaping circuits.

(1) Single differentiating circuit and single integrating circuit.

The system function $G(p)=\tau_2/(1+\tau_1 p)(1+\tau_2 p)$

$$f(t)=(\tau_2/\tau_1)^{1/(1-\tau_2/\tau_1)}\{\exp(t/\tau_2)-\exp(t/\tau_1)\}(\tau_2/\tau_2-\tau_1) \quad . \quad . \quad (11)$$

when normalized to unity maximum amplitude (Elmore 1948).

σ^2 is then a minimum for $\tau_1=\tau_2=\tau=\sqrt{(B/A)}$ and

$$\sigma^2=\sqrt{(AB)}, \quad \frac{e^2}{2}=3{\cdot}7\sqrt{(AB)}. \quad . \quad . \quad . \quad . \quad . \quad (12)$$

A similar response is obtained from using a critically damped tuned circuit with $\tau^2=LC$.

(2) A shorted delay line in conjunction with a single integrating circuit gives optimum signal to noise ratio if the delay line time (for twice the line length) is equal to approximately $1{\cdot}2\tau$ where τ is the integrating time constant, and

$$\tau=\sqrt{(B/A)}.$$

Then

$$\sigma^2=2{\cdot}5\sqrt{(AB)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

(3) A large number of integrating circuits limit the response of a multi-stage amplifier. After a large number of such circuits the response to a pulse $\delta(t)$ tends to the shape

$$f(t)=\exp\{-(t-T_D)^2/\tau^2\} \quad \text{(Elmore 1948)}, \quad . \quad . \quad . \quad (14)$$

where $T_D=n\tau_1$ and $\tau=\tau_1\sqrt{(n)}$.

Where n is the number of stages. When n is large, this is also the response to a function $H(t)$ if this has been differentiated with a time constant small compared with τ. Pulse shaping with a shorted delay line and a large number of integrating circuits will also give approximately the same shape. Then

$$\sigma^2=2{\cdot}5\sqrt{AB} \text{ with } \tau=\sqrt{A/B}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

The optimum case with

$$f(t)=\exp\{-|t-T_D|/\tau\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (15a)$$

with $\tau=\sqrt{A/B}$ gives

$$\sigma^2=2\sqrt{AB}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

These figures differ from those derived by Elmore (1948). Elmore defines a pulse width

$$W^2 = \frac{2\pi \int_0^\infty (t-t_0)^2 f(t)\, dt}{\int_0^\infty f(t)\, dt}, \quad \ldots\ldots\ldots \quad (17)$$

where

$$t_0 = \frac{\int_0^\infty t f(t)\, dt}{\int_0^\infty f(t)\, dt} \quad \ldots\ldots\ldots\ldots \quad (17a)$$

and derives the minimum noise—neglecting the grid noise (*i. e.* putting $A=0$) for a constant value of W. It is of interest to compute W for our various pulse shaping networks when the parameter t_0 has been chosen to make σ^2 a minimum.

For the optimum case $W^2 = 4\pi \frac{A}{B}$.

For the Gaussian case $W^2 = \pi \frac{A}{B}$.

For the RC–RC case $W^2 = 12\pi \frac{A}{B}$.

For the RC–Delay line case $W^2 = 4\pi \frac{A}{B}$.

Thus the Gaussian case has the advantage of giving a smaller pulse width—according to Elmore's definition. This is in fact because there is less of a " tail " to the pulse.

For most purposes it will not be necessary to produce a circuit to shape the pulse to the optimum. The improvement over a Gaussian or RC–Delay line case is only 15 per cent. For some experiments, however, this extra 15 per cent may make all the difference. This arises when low energy protons are to be counted in the presence of intense γ radiation. (Wilson, Collie and Halban 1949). A circuit has been devised (fig. (1)) which will shape the pulses to something which closely approaches the optimum shape. It is necessary, for such a circuit, that only *linear* networks be used ; otherwise, although a pulse $H(t)$ may be shaped to a pulse, $f(t)$ a pulse $2H(t)$ will not be shaped to a pulse $2f(t)$.

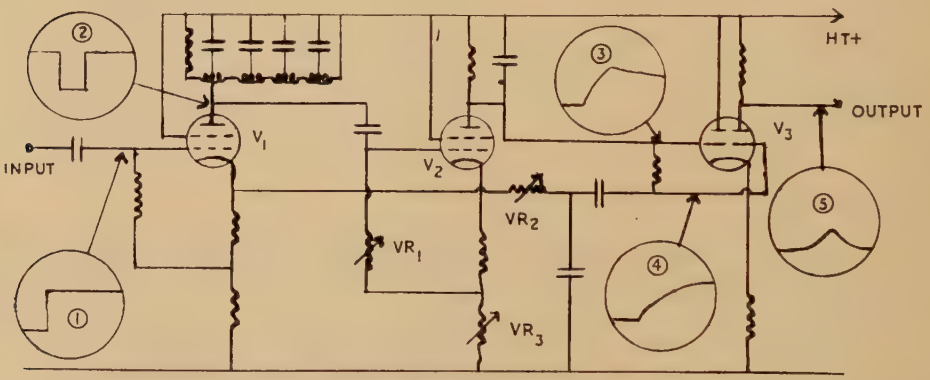
Briefly the operation of the circuit is as follows. The input pulse (1) is split into two parts. One half is converted to a square pulse (2) and then into a nearly linear voltage rise (3). The other half is converted into an exponential rise (4). The first shape is controlled by the delay length and VR1, the second shape by VR2. The relative amplitudes are controlled by the feedback resistor VR3, so that the initial slopes of both waveforms (3) and (4) are the same. The double triode V_3 is connected

so that the difference between the two waveforms is taken as the output. The pulse shape after the peak is exactly the optimum shape, but that before the peak rises too quickly (theoretically it should begin to rise an infinite sime before the peak). With suitable choice of constants

$$\sigma^2 = 2{\cdot}1\sqrt{AB}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

This computation made above of the optimum pulse shaping network assumes that the valve anode and grid noise are independent and constant. If the valve operating conditions are altered, the grid current noises can be reduced, at the expense of the anode current noise, by reducing the HT voltage (and hence the space current) and also by slightly reducing the filament voltage. The minimum noise is given when AB is a minimum (provided that the amplifier time constants are readjusted to the optimum values). If only valve noise and γ background are important this means $(kI_\gamma + I_g)/g_m$ must be a minimum. This is a slight extension of a formula of Gillespie (1948). This assumes that the anode current noise is completely space charge limited, and therefore only holds for a triode.

Fig. 1.



Pulse shaping circuit.

If two different valves are to be compared, then the input capacity alters. A "figure of merit" for a valve in any given application is therefore

$$g_m/C^2(I_g + kI_\gamma). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Where C is the total input capacity of valve and counter (but *not* the feedback capacity due to Miller effect).

Counting of Noise Peaks.

Before it is possible to compare these results with practice it is necessary to consider the rate of counting of noise peaks.

The number of times that the noise waveform passes through a voltage V with positive slope is

$$[\exp\{-V^2/2\overline{V_0^2}\}]\{\tfrac{1}{2} \text{ expected number of zeros of the noise waveform}\}$$

(Rice 1944, 1945, Middleton 1948). . . . . (20)

This is clearly the number of times the discriminator triggers.

The expected number of zeros per second is

$$2\left[\frac{\int_0^\infty f^2 \mathrm{W}(f)\,df}{\int_0^\infty \mathrm{W}(f)\,df}\right]. \qquad \ldots\ldots \quad (21)$$

Where $\mathrm{W}(f)\,df$ is the noise power at frequencies between f and $f+df$.

For the cases considered this approximates to $0{\cdot}6/\tau$ where $\tau=\sqrt{\mathrm{B}/\mathrm{A}}$. (The case of an integrating R–C circuit as the upper frequency cut off, gives a divergent integral; this is discussed by Rice but will not arise in physically realizable circuits.)

For most of our cases $\tau=10^{-6}$ secs.

Then the number of zeros is $0{\cdot}6\times10^6$/second.

Then there will be less than 1 count per minute on noise peaks if $\mathrm{V}/\mathrm{V}_0>6$, *i. e.* if the signal is greater than 6 times the r.m.s. noise.

Practical Amplifier Circuit.

The above analysis has been carried out in order to design a counter-amplifier system to count low energy protons in the presence of intense γ radiation for work on photodisintegration of deuterium (Wilson, Collie and Halban, 1949). The valve chosen for first valve is an American 6AK5 connected as a triode by strapping screen and anode. This reduces the anode current shot noise. The total input capacity of valve and counter is 13 pF. The grid leak is made up only of the valve and insulator leakages and is so high that no noise arises.

The 6AK5 valves used were well aged, and then selected; the grid current was then smaller than any other type tested; the Mullard EF37, although it has a low grid current, gives more anode current noise, and the EC91, or EF91, although they give little anode current noise have a higher input capacity and more grid current noise.

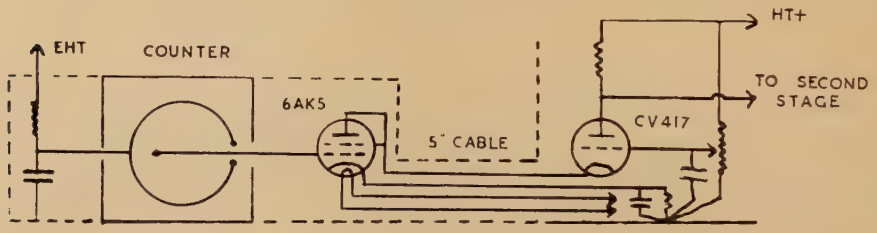
Johnson and Johnson (1936) and Parsegian (1946) recommend pentodes for first valve to avoid the "Miller effect" feedback. This feedback *does not* affect the signal to noise ratio, but it affects the gain and may result in accentuation of second valve noise. This can be obviated by a compensating positive feedback from the second valve anode to the first valve grid (Schultz 1946). We have, however, inserted an extra valve in the "cascode" connection (Hunt and Hickman 1939, Wallman, MacNee and Gadsden 1948). The anode of the first valve feeds directly to the cathode of the second which represents a sufficiently low impedance (100 Ω) to prevent appreciable amplification and feedback by the first valve. Yet the second valve noise is reduced by the high impedance of the first valve in the cathode. This circuit also enables the first valve to be placed away from the subsequent valves—separated by a yard or so of cable (fig. 2).

After amplification by a feedback ring a 100 Ω line is taken to the main amplifier where the pulse shaping networks are situated. A shorted delay line and resistance capacitance integration are normally used. It is possible, with selected valves, to make grid and anode current noise equal with $\tau = 0{\cdot}5 \times 10^{-6}$ secs. with $g_m = 4$ mA/V. Under these conditions, the equivalent number of ion pairs for the noise $= \sigma/c = 300$.

Is is therefore possible to count pulses of ionization of 2000 ions pairs above background, or 70 keV.

When a γ ray source is brought near the high pressure counter used, the value of τ is reduced, to $0{\cdot}3 \times 10^{-6}$ secs. (below which loss of pulse height might occur) and it is only possible to count down to about 90 keV. These values of the background, and the rate of counting of noise pulses (equation (20)) are fully borne out experimentally, confirming that there are no other forms of background than the statistical ones considered here.

Fig. 2.



Amplifier input circuit.

ACKNOWLEDGMENTS.

The thanks of the author are due to Mr. Gillespie and Dr. D. K. C. MacDonald for their advice on noise problems; to Dr. T. W. Chaundy for some mathematical assistance, to Mr. C. H. Collie for suggesting the problem and for his continuous encouragement, and to Mr. G. R. Bishop for his help in the assembly of the apparatus.

REFERENCES.

CAMPBELL, N., 1909, *Proc. Camb. Phil. Soc.*, **15**, 117.
CARSLAW, H. S., and JAEGER, J. C., 1941, *Operational Methods in Applied Mathematics*. O.U.P.
ELMORE, W. C., 1948, *J. App. Phys.*, **18**, 55; *Ibid.*, *Nucleonics*, **2**, 23.
GILLESPIE, A. B., 1948, *A.E.R.E. Report*.
HUNT, F. V. and HICKMAN, R. W., 1939, *Rev. Sci. Inst.*, **10**, 6.
JOHNSON, J. B., 1925, *Phys. Rev.*, **26**, 71.
JOHNSON, E. A., and JOHNSON, A. G., 1936, *Phys. Rev.*, **50**, 175.
MIDDLETON, D. O., 1948, *J. App. Phys.*, **19**, 817.
NORTH, D. O., 1940, *R.C.A. Rev.*, **5**, 244.
PARSEGIAN, V. L., 1946, *Rev. Sci. Inst.*, **17**, 39.
RICE, S. O., 1944, *Bell System Tech. Jnl.*, **23**, 282; 1945, *Ibid.*, **24**, 146.
SCHOTTKY, W. 1918, *Ann. Phys. Lpz.*, **57**, 541.
SCHULTZ, H. L., 1946, *Phys. Rev.*, **69**, 689.
WALLMAN, H., MACNEE, A. B., and GADSDEN, C. B., 1948, *Proc. Inst. Rad. Eng.*, **36**, 700.
WILSON, R., COLLIE, C. H., and HALBAN, H. H., 1949, *Nature*, **163**, 245.

VI. *A Radio Echo Method for the Measurement of the Heights of the Reflecting Points of Meteor Trails.*

By J. A. Clegg and I. A. Davidson,
Physical Laboratories, The University, Manchester*.

[Received September 27, 1949.]

Abstract.

This paper describes a method of finding the heights of occurrence of meteor trails from the radio echoes obtained by broadside reflection on a frequency of 60 Mc/s. The range of the reflecting point from the observing station is measured by the radio pulse technique and its elevation is determined by comparison of the signals received in two aerials mounted at different heights above a horizontal reflecting surface. Results obtained by this method during the Quadrantid shower of 1949 are in good agreement with those of visual observers.

§1. Introduction.

It is well known that a meteor, in its passage through the atmosphere, produces a column of ionization which can be detected by radio echo methods. Such methods have been developed independently by Hey and Stewart (1947) and Lovell (1948) in Great Britain; by McKinley and Millman (1949) Canada; and by Manning, Villard and Peterson (1949) United States.

It has been shown by Hey and Stewart (1947), and by Lovell, Banwell and Clegg (1947), that radio waves of sufficiently high frequency (60 to 75 Mc/s.) undergo specular reflection from the ionized column, the echo being returned from a small element of the trail situated close to the foot of the perpendicular from the station. A study of the durations of such transient echoes presents a number of interesting physical problems, and precise determinations of the heights of the reflecting points is necessary before progress can be made in their solution.

Although the radio echo technique has been used previously to find the heights of occurrence of meteors in the atmosphere, the early methods depended on combined radio and visual observations (Prentice, Lovell and Banwell, 1947, Manning, Villard and Peterson, 1949), and were, moreover, not capable of high accuracy. A more recent method of improved accuracy described by Millman and McKinley (1949) can only be used on low frequencies (30 Mc/s.) where the reflection becomes non-specular.

This paper describes a height finding system which operates on a considerably higher frequency, where the broadside reflection condition

*Communicated by Dr. A. C. B. Lovell.

invariably holds. It is based on the spaced aerial method (Taylor and Westcott, 1946), whereby the range and elevation of the reflecting point from a single observing station are determined by radio measurements alone. The method was used to find the heights of 28 meteors during the Quadrantid shower of 1949, and has since been employed during the daytime showers of 1949 May to July.

§2. Principle of the Method.

The apparatus operates on a frequency of 60 Mc/s. The range of the reflecting point of the meteor trail is measured by the radio pulse method, and its elevation is found by comparing the signals received in two aerials mounted at different heights above a plane, horizontal, perfectly reflecting surface.

For a scattering source of small dimensions at elevation ψ from the station, the ratio, r, of the signals received in two such aerials is equal to the ratio of their effective power gains, $G_1(\psi)$ and $G_2(\psi)$, in this direction, and as these functions are calculable from the known parameters of the system, the relationship between ψ and r can be readily established. For a meteor trail, which subtends a finite angle at the station, the echoes are due to the integrated effect of the radiation scattered from all parts of the ionized column, and the evaluation of ψ in terms of r becomes more difficult. It can, however, be shown as follows, that provided the aerials are not too highly beamed, the amplitude of the echo received in either is proportional to its power gain in the direction of the reflecting point.

Consider first the case of scattering from a trail whose reflecting point lies close to the axis of a wide beam. The overall power gain G, of the combined transmitting and receiving aerial systems, can then be assumed to be constant over the first few Fresnel zones. It has been shown by Lovell and Clegg (1948) that under these circumstances the amplitude of the voltage fluctuations set up in the first stage of the receiver by scattering from a section of the trail lying symmetrically about the reflecting point and subtending an angle 2θ at the station is

$$V = k(\lambda/R)^{3/2}\alpha G\left\{\int_{-v}^{+v} \cos \pi/2v^2 \, dv + i\int_{-v}^{+v} \sin \pi/2v^2 \, dv\right\}. \quad . \quad . \quad (1)$$

Where k is a constant depending on the parameters of the apparatus, λ is the wavelength, R is the range of the reflecting point, α is the electron line density, and $v = 2\theta\sqrt{(R/\lambda)}$. The amplitude of the signal received from the whole trail is obtained by taking the limiting values of the Fresnel integrals as v increases indefinitely.

For accurate elevation finding it is necessary to make observations in directions where the gain of the spaced aerials varies rapidly with θ, so that to derive an expression for the amplitude of the signal received in either it is necessary to include G under the integral sign in (1). The value of V will therefore depend in general on the radiation pattern of the aerial. However,

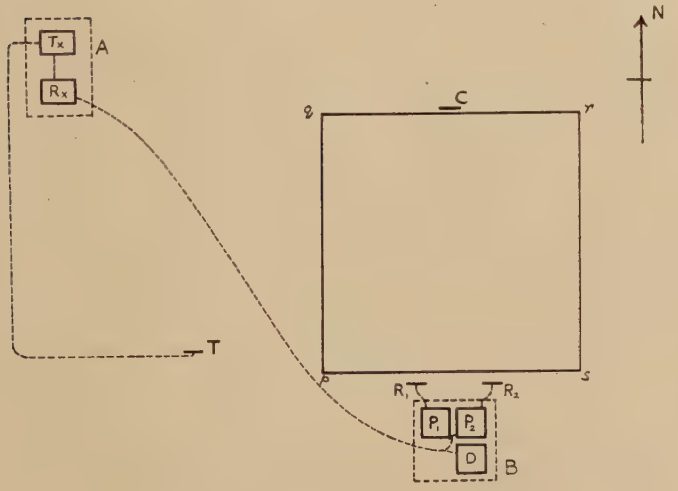
over a small angular range on either side of the reflecting point, G will approximate closely to a linear function of θ, and may be written

$$G_0+\theta\partial G/\partial\theta=G_0+v\partial G/\partial v.$$

Where G_0 is the value of the gain in the direction of the reflecting point. If the width of the beam is sufficiently great in comparison with the angle subtended by the first few Fresnel zones this value may be substituted in (1), which then becomes

$$V=k(\lambda/R)^{3/2}\alpha\left\{G_0\left[\int_{-v}^{+v}\cos(\pi v^2/2)\,dv+i\int_{-v}^{+v}\sin(\pi v^2/2)\,dv\right]\right.$$
$$\left.+\partial G/\partial v\left[\int_{-v}^{+v}v\cos(\pi v^2/2)\,dv+i\int_{-v}^{+v}v\sin(\pi v^2/2)\,dv\right]\right\}.$$

Fig. 1.



Plan of the apparatus (not to scale).

The expressions under the integral signs in the second term of the right-hand side of this equation are odd functions of v, and the integrals vanish identically. The equation therefore reduces to (1), the amplitude of the echo is proportional to G_0, and the ratio of the signals received in the two aerials of a spaced system is equal to the ratio of their gains in this direction.

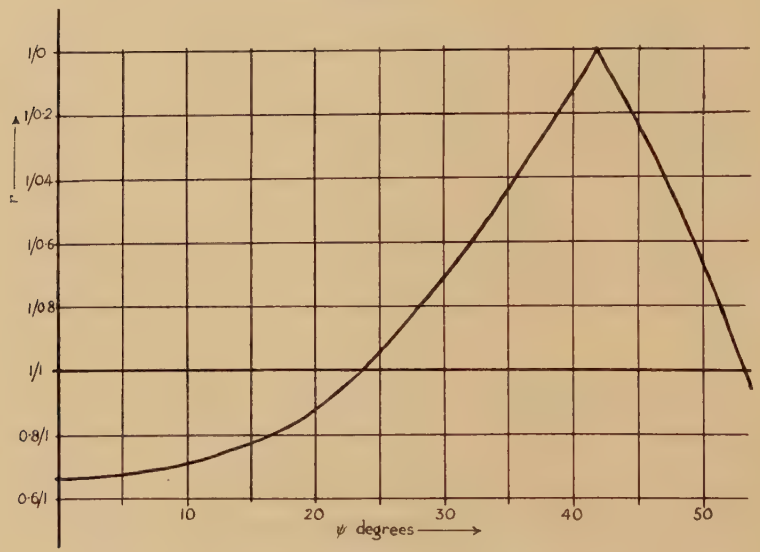
§ 3. Apparatus.

A plan of the apparatus is shown in fig. 1. The hut A contains a transmitter and receiver of conventional design. The transmitter works on a frequency of 60 Mc/s. ($\lambda=5$ metres), and radiates pulses of 8 μ/sec duration and peak power 100 kw. The amplification of the receiver is linear, and its sensitivity is such that a signal of 2 μV. at the input doubles the signal to noise ratio. The transmitting array, T, floodlights the

northern quadrisphere of the sky. It consists of a horizontal folded half-wave aerial at a height of $\frac{5}{8}\lambda$ above the ground, with a reflector adjusted to give minimum backward radiation.

The two receiving aerials R_1 and R_2 are similar to the transmitting array, and are mounted on posts 20 ft. apart at a distance of 200 ft. from the hut A. In front of these aerials is a plane macadam square, *pqrs*, with sides 100 ft. in length which can be covered with wire mesh to form a perfectly reflecting surface. The heights of the receiving arrays above the square can be varied from 3 ft. to 18 ft., but they are normally maintained at $\frac{1}{2}\lambda$ and $\frac{3}{4}\lambda$. The aerials are connected to the two stage preamplifiers P_1 and P_2 housed in the small hut B, and the signals pass from here through a common coaxial cable to the receiver. The

Fig. 2.



Theoretical relation between ratio r and elevation ψ.

preamplifiers are suppressed alternately at a frequency of 25 c/s. by multivibrator pulses supplied from the receiver. The same multivibrator pulses are applied to the x-plates of the A scope display tube, so that the echoes from the two aerials are displayed side by side. Their relative amplitudes are compared visually.

For purposes of calibration a fourth aerial C is mounted at a height of 20 ft. in the centre of the side of the square opposite to the receiving aerials. A duplicate display tube D with its associated time base and power circuits is provided in the hut B. Before each set of readings the aerials R_1 and R_2 are brought to the same height above the reflecting surface and a signal generator is fed into the aerial C. The gains of the preamplifiers are then adjusted until the signals received in R_1 and R_2

are of the same amplitude, and the aerials are then returned to their original heights of $\frac{1}{2}\lambda$ and $\frac{3}{4}\lambda$.

Fig. 2 shows the ratio r of the effective aerial gains plotted as a function of the elevation, ψ. As the aerials fulfil the conditions required by §2, this curve can be used to obtain the elevation of the reflecting point of a meteor trail from the observed value of r.

§4. Limitations and Errors.

(i) *Error in Measurement of the Ratio r.*

With the apparatus in its present form the most serious inaccuracy arises from the uncertainty in the estimation of the ratio r. To find the magnitude of the error involved artificial echoes were displayed on the cathode ray tube, and the apparent visual ratio was compared with that obtained by photographic measurement. The results showed that when the amplitude of the larger echo exceeds that of the noise by a factor of 2·5 or more, the ratio can always be measured to within 1 part in 10, and that for signals differing in amplitude by less than 20 per cent accuracy of 1 part in 20 is possible. For values of ψ greater than 20° the consequent error in the height is never greater than ± 4 per cent.

(ii) *Ambiguity.*

The vertical polar diagram of the higher aerial has two lobes with a zero at 41·8°, so that there are two values of ψ corresponding to each value of r. One of these usually gives an improbable value for the height H, and can be rejected. For a trail at elevations between 33° and 48°, however, there are invariably two possible values of ψ, and the method cannot be used.

(iii) *Error in Range Measurement.*

With the present apparatus ranges can be measured to within ± 2 km. This corresponds to an error in height measurement of ± 1 km.

(iv) *Imperfections in the Reflecting Surface.*

The method depends on the fact that the reflecting surface acts as a plane mirror with a reflecting coefficient of unity. It is well known that for this condition to obtain the surface must behave as a perfect plane reflector over an area comparable with that of the first Fresnel zone. At the lower limit of elevation (20°) the distance of the geometrical reflecting point from the higher aerial is less than 50 ft., but the first Fresnel zone extends outwards to a distance of 180 ft., and a considerable part of its area lies outside the surface *pqrs* in fig. 1. It is therefore important that the ground shall have good reflecting properties for some distance beyond the edges of the square.

For 120 ft. beyond the sides *pq*, *qr*, and *rs*, the earth is at the same level as the prepared surface, and is plane to within $\lambda/16$. The average values given by Burrows (1937) for the dielectric constant and the electrical

conductivity of soil are 10 and 5×10^7 e.s.u. respectively. This implies that for waves of frequencies of 60 Mc/s. its refractive index, n, is equal to $10-i5/3$. Substituting this value in Fresnel's equation we find that for horizontally polarized waves the reflexion coefficient at a glancing angle ψ is given by

$$\rho=\frac{\sin\psi-\sqrt{(10-i\,5/3\cos^2\psi)}}{\sin\psi+\sqrt{(10+i\,5/3\cos^2\psi)}}.$$

In directions within 30° of true north the geometrical reflecting point lies outside the area *pqrs* only when ψ is less than 7°, or when ρ is greater than 0·94. The modification in the vertical polar diagram due to this small departure from perfectly reflecting conditions is inappreciable at elevations greater than 20°. We may therefore assume that over the range of elevations with which we are concerned the curve in fig. 2 is unaffected by the limited area of the prepared surface.

(v) *Coupling between the two receiving aerial systems.*

To measure the degree of coupling between the two receiving aerials, a signal generator was fed into the aerial C (fig. 1), and the signal from the aerial R_2 was permanently suppressed. The amptitude of the signal received in R_1 was measured by a microammeter connected across the output of the receiver. The aerial R_2, which was initially at a height of 3 ft., was raised slowly to a height of 18 ft., and was then disconnected. During this operation there was no perceptible change in the reading of the meter. The coupling between the two aerials must therefore be negligible.

(vi) *Change of aerial impedance with height above the ground.*

The method of calibration described in §3 assumes that the impedance of the receiving aerials is independent of their height above the reflecting surface. To measure the variation in impedance with height a signal was radiated from the aerial C and received in the aerial R_1 in the way described in (v) above. It can readily be shown that if the impedance of R_1 is independent of its height, h, the strength of the signal received will be proportional to $\sin\kappa h$, where κ is a constant. If, however, the mutual impedance between the aerial and its image is appreciable, the signal will show fluctuations in amplitude about the sine curve as h is varied.

On varying h from 3 ft. to 18 ft. these fluctuations were observed, but their amplitude corresponded to less than 0·5 per cent of the total amplitude of the signal. The change in impedance with height is therefore trivial.

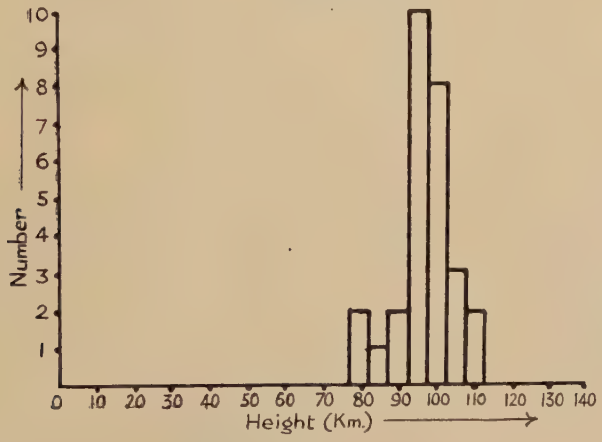
§5. Results obtained during the Quadrantid Shower of 1949.

This method of height measurement was tested during the Quadrantid meteor shower of 1949. The periods of observation were from January 3, $03^h\,08^m$ to $04^h\,40^m$, January 3, $13^h\,00^m$ to $13^h\,49^m$, and January 4,

$03^h\ 08^m$ to $04^h\ 30^m$. At these times the ground was covered with a layer of wet snow to a depth of 5 in. To find the effect of this snow on the reflecting properties of the surface, the experiment described in §4, section (vi) was repeated during the early morning of January 3. No deviation from perfectly reflecting conditions was detected.

The mean echo rate during the periods of observation was 14 per hour. Previous experience with similar apparatus of comparable sensitivity has shown that the rate when no shower is active is two per hour. It may therefore be assumed that the majority of the echoes observed were due to meteors from the Quadrantid radiant. Out of a total of 41 echoes two were at elevations less than 25°, nine gave ambiguous heights, and in two cases there was uncertainty in the measurement of the ratio. The remaining 28 measurements gave heights between 80 km. and 107 km., with a probable error of less than 5 km. in each case. The height distribution is shown in fig. 3.

Fig. 3.



Height distribution for Quadrantid shower, 1949.

§6. Results obtained during the Daytime Showers of 1949, May to July.

The method has recently been used to measure the heights of meteors of the summer daytime streams of 1949. The apparatus was slightly modified to give greater accuracy of range measurement, the pulse recurrence frequency being increased to 150 c/s., and the time base expanded to cover only 250 km. During the period 1949 May to July watches were maintained for a total period of 8·26 hours, and 239 echoes were observed. Of these 166 gave unambiguous heights. The height distribution is shown in fig. 4.

§7. Discussion of Results.

The results obtained by this method show that it is capable of giving

values for the heights of approximately 70 per cent of the total number of meteors observed.

The beginning and end heights obtained visually by Prentice for trails of the Quadrantid meteors are shown in Table I.

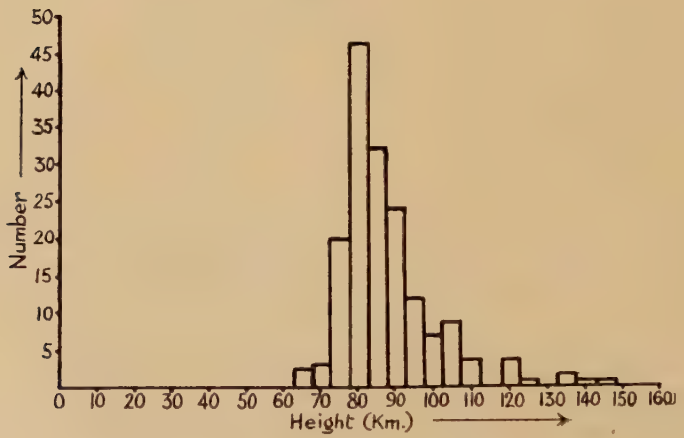
TABLE I.

Heights of Occurrence of Quadrantids obtained from visual measurement.

Visual Magnitude	No. of Meteors Observed	Mean Beginning Height (H_1)	Mean End Height (H_2)
0	1	108 km.	84 km.
2	5	105 km.	87 km.
4	28	103 km.	91 km.
6	140	101 km.	94 km.

Mean Height of Occurrence $\frac{(H_1+H_2)}{2}$ =97·8 km.

Fig. 4.



Height distribution for daytime showers, 1949.

It is evident that the heights shown in this table are confined within somewhat narrower limits than those obtained by the radio echo method, but this is to be expected, since the values shown here represent the means of a relatively large number of observations. The value of the mean height given by Prentice is in good agreement with that obtained by the radio technique.

Although it is not possible to obtain a direct comparison between the results obtained by this and other methods for meteors of the daytime streams, the height distribution shown in fig. 4 is of a form which would be expected for meteors of medium velocity.

§8. Conclusion.

We may conclude, from the results listed above, that the method in its present form is capable of measuring the heights of the reflecting points of meteor trails to within ± 4 km. The discussion in §4 indicates that by far the most serious source of error is the measurement of the ratio of the echoes. This may be eliminated by the addition of automatic photographic recording, which will also reduce the range error. It is estimated that with the addition of this facility, a probable error of ± 1 km. will be attainable.

§9. Acknowledgments.

The authors wish to thank Dr. A. C. B. Lovell, the Director of the Jodrell Bank Experimental Station, for his help and encouragement. The cost of the apparatus, including the construction of the flat reflecting surface, has been met from the grant supplied by the Department of Scientific and Industrial Research for the development of the experimental work at Jodrell Bank. One of the authors (I. A. D.) also wishes to thank the Department of Scientific and Industrial Research for the award of a Maintenance Grant which has enabled him to carry out this work.

References.

Burrows, C. R., 1937, *Bell System Tech. J.*, **16**, 45.
Hey, J. S., and Stewart, G. S., 1947, *Proc. Phys. Soc.*, **59**, 858.
Lovell, A. C. B., 1948, *Phys. Soc. Report on Progress in Physics*, **11**, 415.
Lovell, A. C. B., Banwell, C. J., and Clegg, J. A., 1947, *M. N. Roy. Astr. Soc.*, **107**, 164.
Lovell, A. C. B., and Clegg, J. A., 1948, *Proc. Phys. Soc.*, **60**, 491.
McKinley, D. W. R., and Millman, P. M., 1949, *Proc. I.R.E.*, **37**, 364.
Manning, L. A., Villard, O. G., and Peterson, A. M., 1949, *J. Appl. Phys.*, **20**, 475.
Millman, P. M., and McKinley, D. W. R., 1949, *Sky and Telescope*, **8**, No. 5.
Prentice, J. P. M., Privately communicated.
Prentice, J. P. M., Lovell, A. C. B., and Banwell, C. J., 1947, *M. N. Roy. Astr. Soc.*, **107**, 155.
Taylor, D., and Westcott, C. H., *J. Inst. Elect. Engrs.*, 1946, **93**, part IIIA, 588.

VII. *On the Mass of Charged Particles of the Cosmic Radiation.*

By C. Franzinetti, D. in F.,
H. H. Wills Physical Laboratory, University of Bristol*.

[Received October 31, 1949.]

[Plates I. & II.]

Summary.

The mass of charged particles, produced by cosmic radiation at an altitude of 11,000 ft., and brought to rest in photographic emulsions, has been determinedby 'magnetic-deflection' experiments. The method is applicable to particles, of any type, of which the lifetime is greater than 10^{-10} sec.

The results show that at an altitude of 11,000 feet the particles are composed of different types in the following proportions :—

μ^+-particles, $8{\cdot}5\pm1{\cdot}5$ per cent; μ^--particles, $8\pm2{\cdot}4$ per cent; π^+-particles, $4{\cdot}0\pm1{\cdot}0$ per cent; π^--particles, $5{\cdot}4\pm2{\cdot}0$ per cent; Protons, 48 ± 4 per cent; deuterons, 16 ± 2 per cent; more massive positive particles, 11 ± 2 per cent.

The observations on the tracks of 360 particles give no evidence for the existence of other kinds of particles sufficiently stable to be detected by this method; and they show that if such particles exist, they are very rare.

§ 1. Introduction.

Some time ago it was pointed out by Powell and Rosenblum (1948) that with a suitable arrangement of photographic plates placed in a magnetic field it is possible to measure both the momentum and the residual range of a charged particle, and thence to determine its mass. It was proposed that two plates should be placed parallel to one another and perpendicular to a magnetic field, the two emulsions being face to face. A charged particle leaving one of the emulsions at a small glancing angle, ϕ, and with sufficient energy to reach the opposite emulsion, will be recorded in both plates.

Owing to the magnetic field, the particle will move, in the space between the plates, along a spiral path lying on a circular cylinder of which the radius, ρ, will depend on the component of the momentum of the particle

* Communicated by Professor C. F. Powell.

in the plane perpendicular to the field. From the angle between the directions of motion of the particle, at the points of exit and entry at the two emulsions, ρ can be determined. If the particle stops in the second emulsion, its residual range, r, can also be measured. Inserting the appropriate values of ρ and r in the equations for the momentum, p, and the energy E ; viz. :

$$p=\left(\frac{eZ}{c}\right).\frac{H\rho}{\sin\phi} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and

$$E=f(r, m, Z) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

the mass of the particle, m, can be determined if its charge, Ze, is known.

It was suggested by Powell and Rosenblum that this method would give more precise values for the masses of the π and μ-mesons than can be obtained either by means of grain-counts or by observations on the scattering of the particles. At the time, such improved measurements were of great interest because of their bearing on the nature of the neutral particle emitted during the decay of the π- meson, but shortly afterwards the announcement was made of the artificial production of π-mesons at Berkeley. It seemed clear that these experiments would give much more favourable conditions for the determination of the mass of the particles, and the cosmic-ray experiments were therefore temporarily abandoned.

Interest in the method has been recently revived by reports, from several laboratories, of observations which have been taken to indicate the existence of other types of mesons in addition to the π- and μ-particles. Further, some experimenters (Alichanian *et al.* 1948) have suggested that these other types occur with a frequency comparable with that of the π- and μ-mesons. The question of the existence or non-existence of such particles is of such decisive importance for the development of nuclear physics. It was therefore decided to proceed with the magnetic deflection experiments, both because they appeared to be capable of greater precision than other methods employing photographic plates, and because they have the advantage that any secondary effects associated with the arrest of the particles in the emulsion can be observed. The second of these features is of particular importance because it is reasonable to assume that any other types of mesons must, like the π- and μ-particles, be unstable ; and any positive particles should therefore give distinctive evidence of spontaneous decay.

A further important advantage of the method is that the time of flight of a particle in the apparatus is only of the order of 10^{-10} sec., a period much shorter than that necessary in experiments employing Wilson chambers and counters. It is therefore possible, in principle, to determine the mass of particles, by the photographic method, even although their existence is so transient that they would decay " in flight " in traversing apparatus employing other methods of detection.

The present paper describes an application of the method to the determination of the mass of charged particles observed in photographic plates exposed at the Jungfraujoch High Altitude Research Station.

§ 2. Details of Experimental Procedure.

Apparatus. The experimental arrangement employed is shown schematically in fig. 1(*a*). The plates were carried by a suitable plate-holder, (B), which was placed between the pole-pieces of an electro-magnet. A metal housing, (A), was fitted around the pole-pieces in order to maintain the plate-holder in its correct position during the exposure, and to make it possible to evacuate the space between the plates if it was found to be desirable.

The plates employed were coated with Kodak NT4, or Ilford G5 "nuclear research" emulsions, 50μ, 100μ or 200μ thick, and with linear dimensions $3\frac{1}{4}\times 3\frac{1}{4}$ inches. Emulsions thicker than 200μ have not been found convenient for the investigation, because of the existence of distortion in the processed plates. The most important advantage of using thicker emulsions is that there is an increase in the number of tracks which stop in the emulsion, and which are therefore measurable. The results given in a later paragraph show, however, that the trajectory of a proton with a residual range of 1000μ has a radius of curvature of $\sim$20 cm., in the conditions of the experiment. The corresponding value of the deflection, if the path in the air-gap is $\sim$2 cm., is only 6°, and such a small deflection does not allow a precise measurement of the mass of the particle. The gain in the number of tracks obtained in a given exposure with thick emulsion would therefore be balanced by a reduction in the average value of the precision of the determinations of mass. These considerations do not apply if the observations are confined to the particles of small mass, such as π- or μ-mesons, which are more deflected. It was important, however, for the purposes of the present investigation, to make measurements on particles of all types.

Any small movement of the plates during exposure would have made it very difficult to correlate the two tracks produced by a single particle and would have greatly reduced the accuracy of the measurements. The plates were therefore attached to the holder with "Kodaflat" liquid rubber. The rubber was poured on the frame of the holder in sufficient quantity to make a layer 300μ thick, and was allowed to dry for 12 hours. The plates were then pressed against the adhesive surface, and all possibility of their relative movement was excluded.

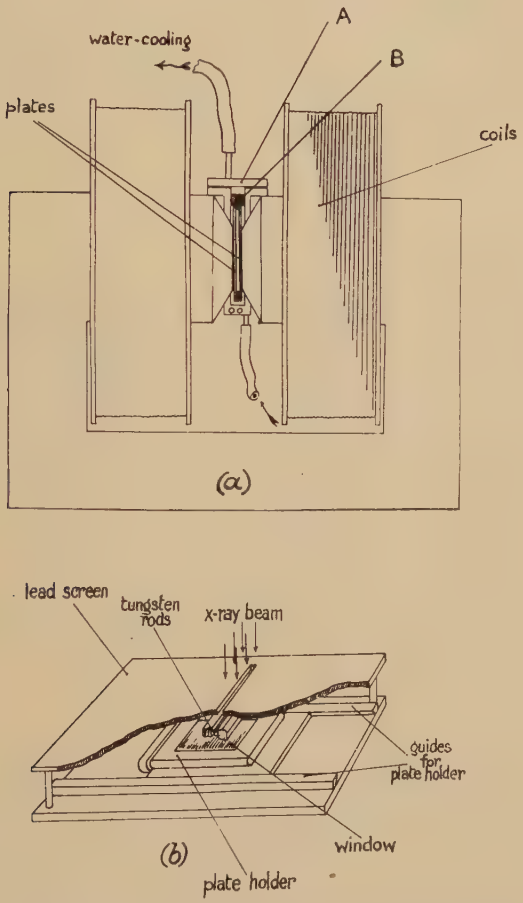
Magnetic Field. A field of 29,180 gauss, in a volume of $\pi\times(3{\cdot}5)^2\times 0{\cdot}7$ cm.^{3}, was maintained by an electro-magnet, the necessary power being 600 watts, at a current of 8·5 amps. The field was constant during the exposure to within 300 gauss, apart from "edge effects", which were only appreciable in the region extending 4 mm. from the edges of the pole-pieces. The variation of the field with the current is shown in Table I., and it will be seen that it is very small for currents between 8 and 9 amps. During the exposure, the changes in current never exceeded 0·25 amps.

TABLE I.

Variation of field-strength with magnetizing current.

I (amps.)	8·0	8·2	8·4	8·5	8·6	8·8	9·0
H (gauss)	28,900	29,020	29,140	29,200	29,250	29,340	29,440

Fig. 1.



Experimental arrangements : (*a*) during the exposure,
(*b*) X-ray marking device.

The field was determined by comparison with a known field produced by a permanent magnet, kindly made available by Dr. S. Rosenblum and Mr. Tsai of Bellevue Laboratory in Paris, using a fluxmeter of high precision. A cooling system maintained the plates below 20° C.

Correlation of Pairs of Tracks. A precise knowledge of the geometrical relationship between the two plates during exposure was required in

order to correlate the pairs of tracks produced by each particle, and this was obtained by the following method :—

The holder containing the plates was exposed to an X-ray beam, passing perpendicular to the plane of the emulsion ; (see fig. 1(*b*)). Between the plates and the X-ray source was a screen of lead a few mm. thick, and provided with a narrow slit 8 cm. long and a few microns wide. With this arrangement it was possible to " print " a grid of lines on both the plates, as indicated in fig. 2, in such a way that corresponding points on the grids represented the same x, y, coordinates. The edges of the slit consisted of two cylinders of tungsten, 2 mm. in diameter and it was made sufficiently narrow by pressing a piece of thin paper between the two cylinders during assembly. The black lines produced in the processed plate by X-ray exposure had a width of 20μ, or even less.

Fig. 2.

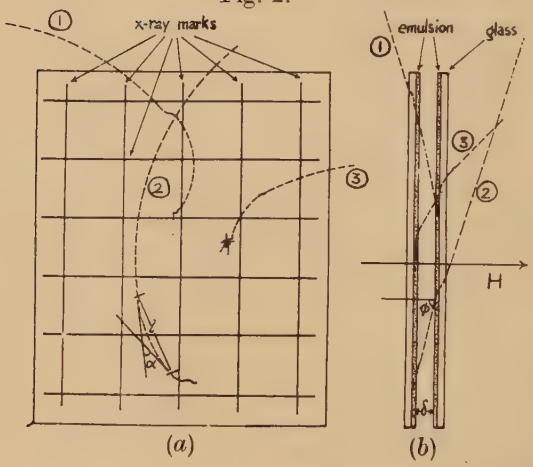
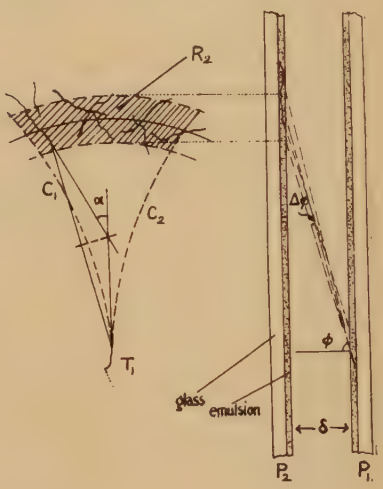


Fig. 3.



The rectangular grid of lines on the two plates greatly facilitated the correlation of the tracks. The two plates of each pair were first scanned, and maps were constructed showing the track of all particles crossing the surface of an emulsion, and of those which stopped in one of them. From the position, direction, and angle of dip, ϕ, of a track, T_1, which stopped in one emulsion, (see fig. 3) it was possible to determine a region, R_2, in the opposite plate, in which the continuation of the track should be found. If the density of the tracks in P_2 is low enough, only the true continuation of T_1 will have the correct ionization, angle of dip, and direction, consistent with it being the partner of track T_1. The probability that the track produced by another particle will, by chance, satisfy the same requirements is discussed in §4, and is shown to be very small.

Processing of the Emulsions. Great care was taken in processing the plates in order to reduce the distortions of the emulsion to a minimum. The following procedures have been found satisfactory :

Chemicals	Temperature	Time 50 μ	100 μ	200 μ
For Kodak NT4.				
Developer (1 part D19b, 2 ,, H_2O.)	18° C.	30m.	40m.	60m.
Hardening Stopping Bath	,,	10m.	15m.	20m.
Fixer (Hypo 30 per cent Solution in weight)	,,	3/2 of clearing time.		
Washing in water	,,	3h.	4h.	5h.
For Ilford G5.				
Developer	3° C.	—	—	35m.
16 cc. Azol Dev. 10 cc. 1 per cent K Br. Sol. 800 cc. H_2O.	18° C.	—	—	200m.
Hardening Stopping Bath, Fixer	As for Kodak NT4.			

§ 3. Method of Calculation of Mass.

According to Camerini and Lattes (unpublished), the range-energy relationship for protons, in a photographic emulsion, can be represented by a simple power law for values of the range greater than 200μ :

$$E = KR^n,$$

where E is the energy of a proton of range, R. Further, the ranges R(v) and $r(v)$ of particles having the same initial velocity, v, but with masses M and m, and charges Ze and ze, respectively, are related by the equation

$$R(v) = r(v) \cdot M/m(z/Z)^2 - C,$$

C being constant*. Since C is negligible compared with the ranges of particles considered in the present experiment, it may be neglected. It follows that the relation between the energy ϵ and the range, r, of a particle of mass m, and charge ze, is given by

$$\epsilon = K(m/M)^{1-n} z^{2n} r^n, \quad \ldots\ldots\ldots \quad (3)$$

M being the mass of the proton, and e the elementary electronic charge.

The equation (3) is the explicit form of the equation (2). Solving the simultaneous equations (1) and (3), we obtain

$$\frac{m}{Z^\alpha} = A \left(\frac{H\rho}{\sin \phi} \right)^{\frac{2}{2-n}} r^{\frac{n}{2-n}}, \quad \ldots\ldots\ldots \quad (4)$$

* (The constant C takes account of the capture and loss of electrons in the last part of the range, a process which affects differently particles of different charge. See Bethe and Livingston (1937).)

where $\alpha = \frac{2(1-n)}{2-n}$ and $A = \left\{\frac{e^2}{c^2} M^{1-n}/2K\right\}^{\frac{1}{2-n}}$.

For ranges shorter than 200μ, the range-energy relation given above is no longer valid, and the function can therefore only be evaluated numerically from the experimental points. If the logarithm of the residual range of a particle is plotted against the logarithm of its measured radius of curvature, it follows, from the equation (4), that the particles with equal masses should be represented by points lying on straight lines, of which the slope, ψ, is given by

$$\tan \psi = n/2.$$

There will be a departure of the observations from this linear relationship, however, in the case of particles of short range.

§4. Sources of Errors.

The most important errors in the experiment are due to : (i) distortion of gelatine ; (ii) scattering in the air-gap ; (iii) scattering in the surface layers of the emulsion ; and (iv) spurious coincidences.

(i) *Distortion of gelatine.*

Serious distortions of the emulsion extending over large areas, which can change the coordinates of a point by more than 10μ, have been found to be confined to regions near the edges of the plate. The affected area extends inwards from the edges of the plate for 2 mm. in the case of emulsions 50μ thick ; and for 1 cm. for 200μ emulsions. On the other hand, local distortions, which may considerably alter the direction of a track, sometimes occur if special care is not taken in processing. Local distortions are very often accompanied by differences in the general background of grains ; and they may sometimes be detected, with the naked eye, in the form of manifest imperfections in the smooth surface of the emulsion.

(ii) *Scattering in the air-gap.*

In crossing the gap between the plates, the particle may be scattered by the gas filling it. The influence of such effects on the measurement of the curvature of the tracks of particles in magnetic deflection experiments, has been discussed by Bethe (1948). His calculations, when applied to our particular experimental conditions, lead to the result given in Table II., which shows the values of $\frac{\rho_H}{\rho_S}$, the ratio of the radius of curvature of a track due to the magnetic field, to the average apparent radius of curvature due to scattering. $\frac{\rho_H}{\rho_S} = \frac{\delta_S}{\delta_H}$, where δ_S and δ_H represent the angular deviations due to scattering and the magnetic field, respectively.

(ii) *Scattering in the surface layers of the emulsion.*

The determination of the direction of motion of a particle, at the exit from one emulsion and at entry into the other, is subject to errors as a

result both of scattering by the atoms of the emulsion, and of the displacement of the individual grains in the track from the true line of the trajectory. The second of these effects has been studied by Goldschmidt *et al.* (1948), in experiments with C2 Ilford plates, and has been found to be less than three per cent of the real scattering for tracks of range less than 2000 μ. The error from "spurious scattering" can therefore be regarded as negligible.

The errors due to scattering in the surface layers in the emulsion depend on the method of defining the direction of the track. If we choose to define this as the direction of the chord in the last few microns as the track approaches the surface of the emulsion, the average error can be evaluated :

From the formulæ given by Rossi and Greisen (1941), and using their notation we easily obtain for the angle ξ between chord and trajectory

$$\langle \xi^2 \rangle = 4/3(t/w^2),$$

TABLE II.

Values of $\rho_H/\rho_S = \delta\,\delta_S/\delta_H$; H=29,200 ; l=1 cm.

Residual range	Mass of Particle		
	200 m_e	300 m_e	2000 m_e
μ	(per cent)	(per cent)	(per cent)
100	2·7	2·9	4·3
200	2·3	2·55	4·0
400	2·0	2·15	3·4
600	1·85	2·03	3·0
800	1·7	1·9	2·9
1000	1·0	1·8	2·7

The calculations have been made for a particle of charge e, and masses $\sim$200 m_e, 300 m_e and 2000 m_e for a path length in the air-gap of 1 cm. The values of ρ_H/ρ_S, when multiplied by $2/(2-n)$ (about 1·4 in the present instance) give the error in the mass due to the scattering in the air.

where t is the length of the chord in radiation units, and w is a function of the momentum and velocity. On the other hand, for the total angular deviation θ of a track from its original direction, after a path, t, we have

$$\langle \theta^2 \rangle = 2t/w^2 = 1{\cdot}5\langle \xi^2 \rangle .$$

It follows that $\langle \xi^2 \rangle$ is easily evaluated from experimental values of $\langle \theta^2 \rangle$ (Goldschmidt *et al.* 1948).

A simple and more direct method of estimating the importance of all the above sources of error is to make an exposure without a magnetic field. The measured "deflections" are then due to scattering or distortion. An examination of a plate containing a total of 50 tracks (10 mesons and 40 protons) has shown that the errors in the angular measurements are distributed in a gaussian curve, the half-width of which is of the order of 1·4°.

(iv) *Spurious coincidences.*

The number of " spurious coincidences "—the number of cases in which two unrelated tracks are attributed to the same particle—depends on the criterion adopted for the recognition of true pairs—and the method employed has been described in §2. Consider a track T_1 in P_1, and let its angle of " dip " be $\phi_1 \pm \Delta\phi$, its residual range, r_1, and its direction α_1. The region on P_2 in which a search for the partner track will be made, is the area R_2, shown in fig. 3 ; this area is limited by two circles of radius $\delta \tan(\phi_1 - \Delta\phi)$ and $\delta \tan(\phi_1 + \Delta\phi)$ on the one hand ; and by the circular paths C_1 and C_2, corresponding to the maximum deflection expected for the particles with mass greater than 100 m.e., on the other.

Suppose now that the tracks in P_2 are isotropically distributed. This assumption is only approximately correct, but it can hardly lead to serious errors. Suppose, further, that the tracks occur with a frequency per unit area equal to ν_2. The number of " fortuitous coincidences " will then be given by the number of tracks contained in R_2, crossing the surface of the emulsion, with the correct angle of dip, $\phi \pm \Delta\phi$, and correct direction, $\alpha_2 \pm \Delta\alpha$. If $\Delta\phi$ is small, the required value is given to a sufficient approximation by

$$n = \frac{2\nu}{\pi} \cdot \frac{\Delta\alpha(\Delta\phi)^2 \cdot \delta^3}{\rho_{\min}} \cdot \frac{\tan^3\phi}{\cos\phi}.$$

For N tracks, the total number of " coincidences ", C, is,

$$C = \sum_1^N n.$$

If the N tracks are isotropically distributed within the angles ϕ_1 and ϕ_2, we then obtain

$$C = \frac{2N\nu \cdot \Delta\alpha \cdot (\Delta\phi)^2\delta^3}{\pi\rho_{\min}} \cdot \left[\tfrac{1}{3}\tan^3\phi - \tan\phi + \phi\right]_{\varphi_1}^{\varphi_2}.$$

In the conditions of our experiment :

$$\Delta\alpha = 1° = 0{\cdot}017 \text{ radians},$$
$$\Delta\phi = 2° = 0{\cdot}034 \text{ radians},$$
$$\delta = 0{\cdot}3 \text{ cm.},$$
$$\rho_{\min} = 2 \text{ cm.}$$

Considering only particles having a path, l, in air between 1 and 4 cm., we find, from Table III., that $C \sim 10^{-4}\nu N$.

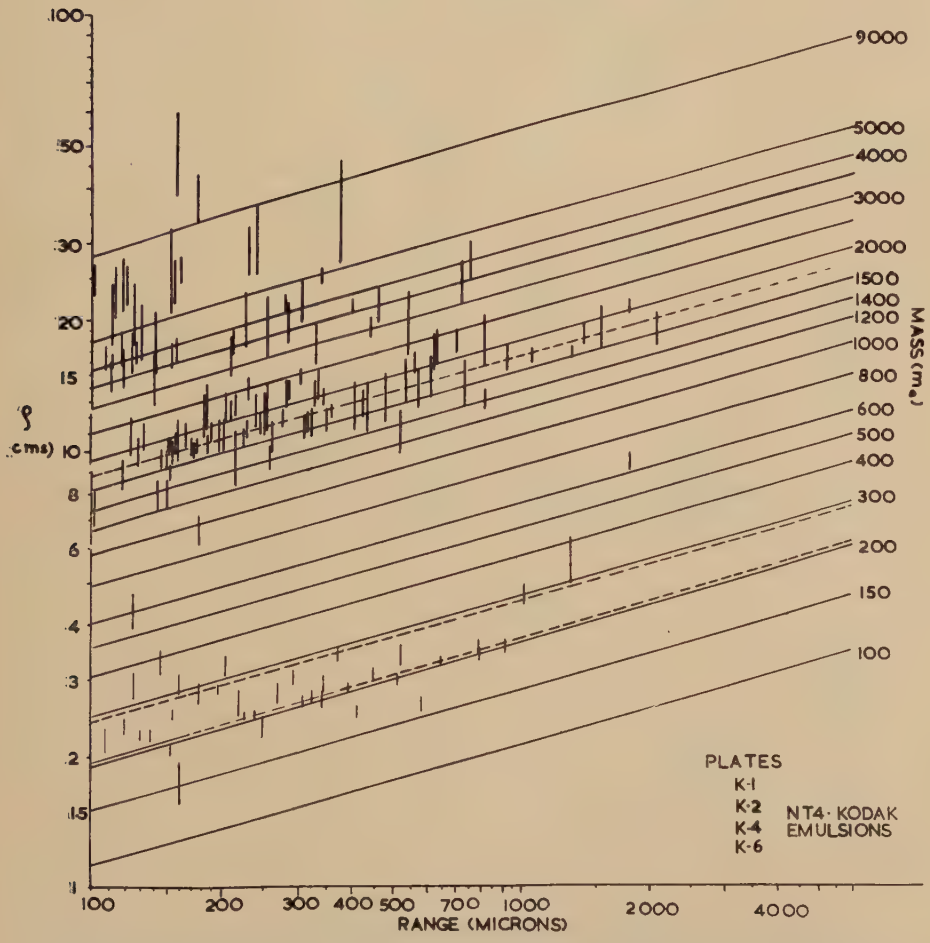
Table III.

Values of $\Phi = [\frac{1}{3}\tan^3\phi - \tan\phi + \phi]$					
l (cm.)	1·0	1·5	2·0	3·0	4·0
ϕ (rad)	1·27	1·37	1·42	1·47	1·50
Φ (ϕ)	10·4	35·2	92·7	324·5	788·0

If $\nu < 100$ cm.$^{-2}$ the number of fortuitous coincidences will be less than 1 per cent.

In this calculation, it has been assumed that the correlation of the events has been made only on the basis of geometrical considerations, whereas in actual fact, the observation of a close similarity in the grain-density of the two tracks of a pair has always been considered an essential additional criterion. We have assumed, however, an isotropic distribution of the tracks which (see fig. 7) does not, in fact, exist. Approximate

Fig. 4.



Logarithm of radius of curvature ρ (cm.) versus logarithm of residual range r. The lines represent the function with slight corrections to allow for deviations of the range-energy relation from the power law for short ranges.

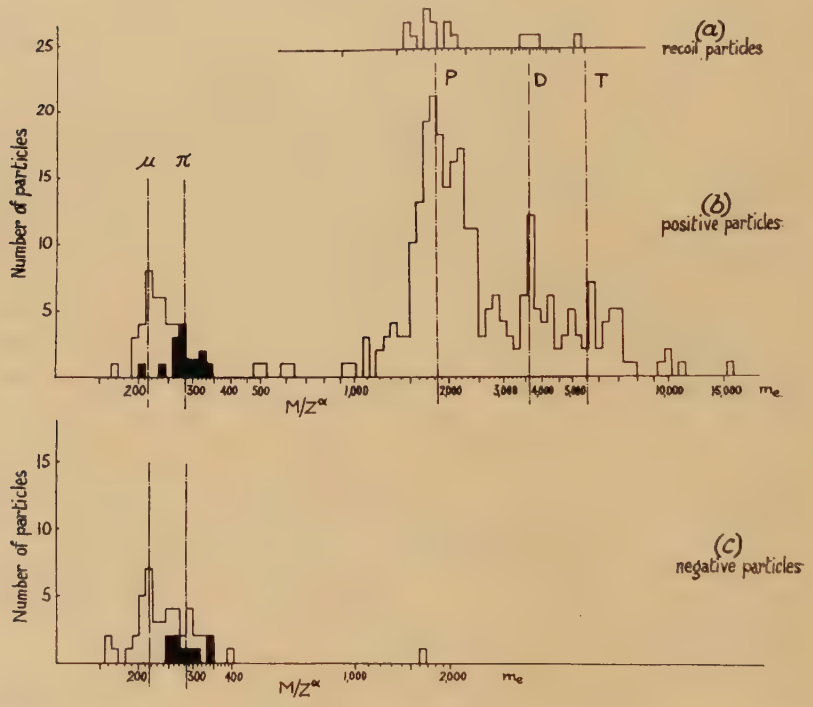
estimates suggest that the two effects compensate one another, and it is reasonable to suppose that the calculations give a satisfactory indication of the order of magnitude of the number of spurious coincidences.

§ 5. Results.

In fig. 4 the results of measurements on the tracks of 150 particles, observed in Kodak NT4 emulsions, have been plotted. The probable

error in the measurement of the radius of curvature, calculated in each case as discussed in § 4 is indicated for each point. For the sake of clarity, the error in range, which is estimated to be smaller than 5 per cent in every case, has not been shown. A similar graph has been constructed for the other 210 measurements, made with Ilford G5 plates, for which the range-energy relation is slightly different, and the results are essentially similar. It will be seen that the results clearly show the division of the particles into groups corresponding to mesons, protons and heavier

Fig. 5.



Mass distribution of 360 recorded particles. The dotted lines correspond to mass values of μ-mesons (215 m_e), π-mesons (285 m_e), protons, deuterons, tritons. It follows from equation (4) that a particle of charge Ze appears (using this method) to have a mass of $m\,Z^{\alpha}$, α being 0·59 for Kodak NT4, and 0·56 for Ilford G5. $\pi+$ and σ_- mesons are indicated in black.

particles, the most prominent group being that due to protons. The linear distribution, anticipated from equation (4), for particles of a given mass and of range greater than 200 μ, is clearly displayed.

The results of measurements of the mass of 359 particles are shown graphically in fig. 5, and details of the measurements on the group of particles with masses less than 400 m_e, including the secondary phenomena observed at the end of their range, are given in Table IV.

TABLE IV.

Analysis of Mesons—Mass Measurements.

Plates Sensitive to Minimum Ionization Tracks.

The mesons are classified phenomenologically, according to the secondary effects observed at the end of their range, into π, σ, ρ (Powell, 1949). A μ-meson is one which is *observed* to originate by the decay of a π-particle in the emulsion.

No.		Sign	α (Deg.)	ρ (cm.)	Residual Range	Mass	Probable error in Mass	Observations
K.O.	12	−	13·0	1·94	105	220	20	ρ electron decay
K.1.	11	+	17·2	2·18	108	240	30	ρ electron decay
	21	−	27·5	2·55	415	178	20	ρ
	29	+	29·5	2·94	308	215	10	ρ
	26	+	23·0	3·63	921	215	18	ρ
	76	+	19·0	2·24	130	235	15	ρ electron decay
	90	+	21·0	2·80	178	285	35	π–μ decay
	110	−	24·1	2·22	137	230	15	ρ (slow electron)
	158	−	30·2	2·71	344	210	18	ρ electron decay
	164	+	37·5	2·56	258	220	10	μ
	72	−	42·0	2·96	126	350	50	σ (5 branch star)
	55	−	26·7	2·76	327	216	10	ρ (blob)?
	125	−	61·5	3·63	650	215	7	ρ (slow electron)
	142	+	51·5	3·1	395	206	10	μ (electron decay)
	39	−	35·5	3·84	249	390	30	ρ (blob)
K.2.	24	−	22·5	3·09	295	270	10	σ (3 branch star)
	89	−	25·0	2·54	585	158	25	ρ
	75	+	21·0	2·3	57	240	30	π
I.7.	2	+	15·8	3·55	388	280	20	π
	6	+	8·0	3·7	286	330	30	π
	96	+	33·0	2·5	305	190	30	ρ electron decay
	118	+	15·0	2·7	315	120	18	ρ electron decay
	133	+	19·0	1·9	220	155	20	ρ electron decay
	141	−	16·0	3·94	690	250	20	σ (1 branch star)
	156	−	9·3	3·8	233	320	30	ρ slow electron
	186	+	29·0	2·7	310	212	20	μ electron decay
	10	+	10·0	2·85	335	220	20	μ electron decay
	35	−	23·7	1·9	145	205	20	ρ
	54	−	44·4	1·85	101	220	7	ρ (blob)?
	121	−	10·0	3·33	357	265	30	ρ (blob)
	134	+	12·0	2·5	124	310	30	π
	190	−	13·0	2·35	235	200	40	ρ (blob)?
	191	+	11·0	2·7	206	250	20	ρ
	173	+	12·0	2·94	335	225	15	ρ electron decay
	176	+	8·0	2·88	290	235	20	ρ electron decay
	177	−	19·6	2·88	355	220	12	ρ electron decay
I.8.	64	−	31·6	2·9	298	235	10	ρ slow electron
	76	+	13·4	2·95	165	320	25	π
	81	+	13·9	2·58	270	210	25	μ electron decay
	85	−	13·5	3·55	430	265	25	σ (2 br. star)
	93	+	24·4	4·44	1530	190	15	ρ electron decay
	91	+	10·0	2·6	241	230	20	ρ electron decay
	99	−	12·5	1·9	200	165	25	ρ

TABLE IV. (*Cont.*)

No.		Sign	α Deg.)	ρ (cm.)	Residual Range	Mass	Probable error in Mass	Observations
	154	—	22·5	2·78	396	195	10	ρ (blob? and slow electron)
	164	+	14·0	2·14	160	170	25	ρ electron decay
	174	+	16·5	3·0	370	225	15	ρ
	171	+	14·0	2·05	510	250	30	μ
	140	—	19·0	1·7	120	180	50	ρ
	226	—	12·0	2·55	140	292	50	σ (1 heavy trk.)
I.9.	29	—	17·0	2·75	202	260	20	ρ
	46	+	15·0	4·45	714	280	20	π
	47	+	17·5	2·6	200	245	20	μ (incomplete)
	52	+	14·4	3·0	400	215	5	ρ electron decay
	65	+	25·5	2·45	295	192	10	ρ electron decay
	74	—	9·0	4·0	450	300	20	ρ
	79	+	10·5	2·35	236	200	40	π
	95	—	17·1	4·0	508	290	20	ρ slow electron
	105	—	9·0	3·0	108	350	40	σ (1 heavy br.)
	121	—	19·0	3·33	392	245	20	ρ (blob)?
	125	—	19·8	4·0	484	295	30	ρ (blob)
	3	—	14·5	3·28	496	218	25	ρ (blob)?
	8	+	20·9	1·95	135	207	15	ρ electron decay
	16	+	7·0	3·7	460	260	20	ρ electron decay
	30	—	13·5	4·5	902	250	40	ρ electron decay
	95	+	9·0	4·75	1700	200	20	ρ electron decay
	84	—	9·0	3·2	227	310	30	σ (4 branch star)
			Plates not Sensitive to Minimum Ionization Tracks.					
K.4.	26	—	14·0	2·34	120	262	25	σ (4 branch star)
	63	—	13·7	1·75	160	160	30	ρ
	71	—	12·0	2·33	250	195	20	ρ
	30	—	24·5	2·99	512	207	15	ρ (slow electron)
	31	—	16·5	2·81	267	240	25	ρ
	38	+	91·6	2·5	228	220	10	ρ
	41	—	22·5	2·3	150	230	15	ρ
	44	+	26·0	3·65	800	225	15	ρ
	91	—	13·4	3·48	800	208	15	ρ
	28	—	21·0	3·48	374	282	15	ρ
	107	—	10·8	2·66	220	245	30	σ (3 branch star)
	47	+	20·0	2·92	346	230	20	ρ
	111	+	9·5	5·7	1300	340	80	π
	100	+	14·9	4·72	1030	292	20	π
K.6.	68	—	21·5	3·26	205	330	30	ρ
	70	+	17·0	2·88	197	283	7	π
	114	—	11·0	2·05	85	250	50	ρ
	92	+	31·0	1·75	45	258	20	π
	97	+	33 0	2·17	91	262	13	π
	19	+	17·0	2·5	160	260	15	π
	26	—	28·5	3·1	450	228	15	ρ
	45	—	30·0	2·07	153	200	17	ρ
	53	+	12·5	3·45	523	235	16	μ
	57	—	11·0	2·94	160	315	40	ρ
	109	—	26·4	2·5	239	217	10	ρ

The observed distributions of the values of the mass of the particles, taken separately according to their sign, are shown in figs. 5(*b*) and (*c*).

The distribution in mass of the positively charged particles (fig. 5(*b*)) shows :

(*a*) A peak due to particles with masses in the interval from 160 to 250 m_e, the mean value being $217 \pm 4\ m_e$*. About 80 per cent of these particles decay with the emission of an electron, of which the track can be distinguished. Six of them are μ-mesons created by the decay of π-particles in the emulsion.

(*b*) A peak, just resolved from the first, and due to particles with masses between 240 and 340 m_e, the mean value being $281 \pm 10\ m_e$. All these particles decay at the end of their range with the emission of a μ-particle.

(*c*) A more prominent group, corresponding to the particles with a mean mass of $1870\ m_e$, $\pm 27\ m_e$. None of these particles give any indication of inter-acting with matter at the end of their range, nor do they produce any observed charged particles by spontaneous decay.

(*d*) A peak of lower intensity can be seen at about $3700\ m_e$, and an indication of a broad distribution due to particles of a greater mass. None of these particles give any indication of inter-acting with matter, nor of undergoing spontaneous decay.

The distribution of values of the mass of negative particles (fig. 5(*c*)) shows :

(*e*) A peak between 160 and $250\ m_e$, and a broader distribution extending up to $400\ m_e$. Of the particles with masses between 160 and $250\ m_e$, only four show electron decay ; and five others give rise to slow electrons originating at the end of their range. Of the particles between 250 and $400\ m_e$, 9 produce nuclear disintegrations, and the others stop without leading to the production of any secondary charged particles for which the evidence is unambiguous. In ten cases there is an unusual thickening at the end of the track, see Pl. II.(*b*), due to an assembly of two or three grains which might be attributed to a secondary particle of very short range ; but in all cases except three it has been impossible to decide whether this " blob " was produced by a slow electron or by a heavy particle. In the three cases the " blob " was definitely too thick to have been produced by an electron.

(*f*) In fifteen cases, a particle which crosses the air-gap and is brought to rest in one emulsion, originates in the emulsion of the other plate. The complete track therefore has two " ends ". The direction of motion of the particle, in events of this type, can be determined by the increased grain-density in the track as the particle slows down, and the more numerous deviations due to Coulomb scattering towards the end of the range. The particles of this type are all positively charged and the distribution in the observed values of their mass is shown in fig. 5(*a*).

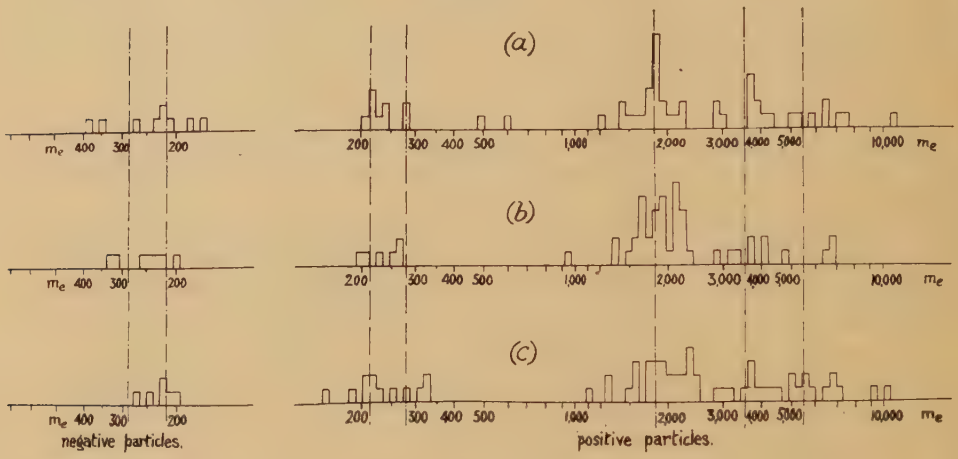
---

* The errors, σ, are calculated from the residuals, v_i, according to the formula

$$\sigma = \sqrt{\frac{\Sigma v_i^2}{n(n-1)}}.$$

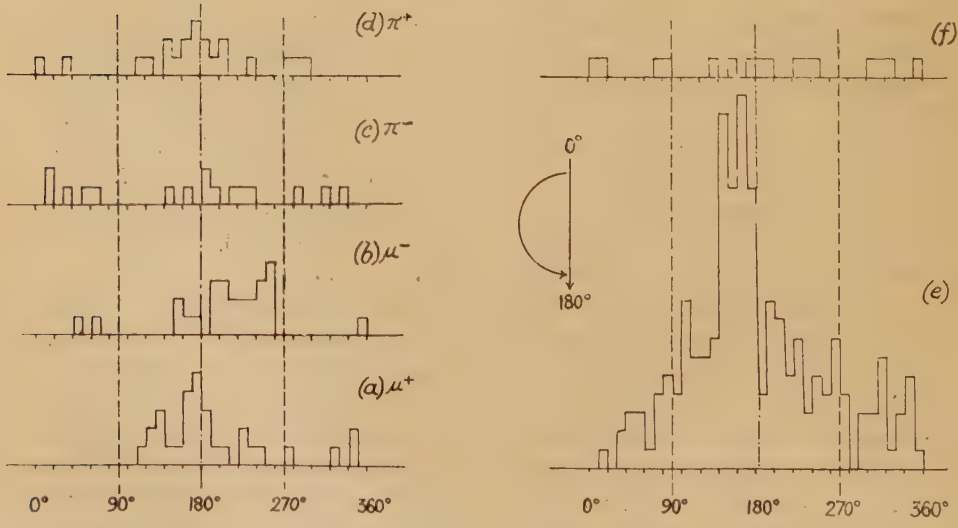
(*g*) The angular distribution, relative to the vertical, of the directions of motion of all particles on which measurements have been made, at their points of entry into the first emulsion through which they pass, are shown in fig. 7. In this figure, the particles have been grouped together according to their masses and the signs of their charge. In

Fig. 6.



Effects of distortion of the gelatine on the precision of the measurements.
(*a*) 61 tracks from 2 pairs of 50 μ thick emulsions (25 days exposure).
(*b*) 56 tracks from 1 pair of 100 μ thick emulsions (15 days exposure).
(*c*) 69 tracks from 1 pair of 200 μ thick emulsions (10 days exposure).

Fig. 7.



Angular distribution of: (*a*) μ^+; (*b*) μ^-; (*c*) π^-; (*d*) π^+; (*e*) recoil particles; (*f*) protons, deuterons and heavier particles. This histogram also contains some tracks which were not in a suitable position for being measured, although their identity could be established with certainty.

most cases there is a tendency for the directions of motion to be grouped round the vertical, but to be deflected from it in a sense depending on the sign of the charge.

(*h*) There is a gradual reduction in precision as the thickness of the emulsion employed is increased—see fig. 6.

§6. Interpretation of Results.

From the evidence given in sections (*a*) to (*g*) above it is concluded that :

(α) The positive particles corresponding to group (*a*) are μ^+-mesons created by the decay of π^+-particles. In those cases in which the μ^+-particle has an energy greater than 4·2 MeV., we must assume that the parent π^+-particle decayed whilst in motion. The average value of the mass of these particles is $217 \pm 4\ m_e$. We therefore identify them with the mesons of the penetrating component of the cosmic radiation, of which the mass, as determined at Berkeley, is $215 \pm 4\ m_e$. 80 per cent of these particles decay with the emission of an electron. In view of the fact that "thin" tracks inclined steeply to the plane of the emulsion are difficult to detect, and since there is an isotropic distribution of the directions of motion of the decay electrons, we may assume that we fail to detect 20 per cent of the electron tracks, of which the specific ionization has the minimum value, in the conditions of the experiment ; but that, in fact, all the μ^+-mesons decay with the emission of an electron.

(β) The particles described in group (*b*) are obviously π^+-particles. From the observed value of the mass, $281 \pm 7\ m_e$, they can be identified with those produced artificially.

(γ) The negative particles described in group (*e*) are made up of μ^-- and π^--particles. Nine of them produce observed disintegrations, and of these the average value of the mass is $288 \pm 13\ m_e$. They can therefore be identified as π^--particles. The average mass of the other 38 negative particles, only four of which appear to decay with the emission of an electron, is $236 \pm 8\ m_e$. It is reasonable to attribute this value, intermediate between the masses of the π and μ particles, to the presence among the μ^--mesons, of some π^--particles which produce no visible disintegrations. The observed value of the average mass suggests that 28 ± 4 of these mesons are μ^--particles, and 10 ± 4 are π^--particles.

(δ) The group with apparent masses between 1100 and 2400 m_e., which contains 48 per cent of all tracks measured, is formed by positive particles which stop in the emulsion without showing any inter-action with matter, nor any evidence of spontaneous decay. Their average mass is $1870 \pm 27\ m_e$, and they may be identified with confidence as protons.

(ϵ) The group between 2400 and 4400 m_e, of smaller intensity than the protons and less clearly resolved, is also made up of positive particles which produce no secondary effects ; they are therefore regarded as deuterons.

No accurate discrimination of particles of different types is possible in the region of the spectrum beyond 5500 m_e. This is due not only to the small deviations suffered by heavy particles, but also to the fact that

their charge, which may be different from $|e|$, is unknown. We therefore attribute this group to a mixture of tritons and α-particles, which cannot be resolved in the conditions of our experiment.

(ζ) Of the seven particles which appear to have an intermediate value of the mass, six are positively charged, and one appears to be negatively charged. We must expect that a certain proportion of the different particles which we record will be scattered through an angle of a few degrees or more, in passing through the gas between the plates, as the result of elastic collisions with oxygen or nitrogen nuclei. Such a deviation in the trajectory will usually lead us to reject the particle from among those which give tracks acceptable for measurement, for the directions of motion—at the points of exit from and entry into the two emulsions—will not in general lie on a circle. In a small proportion of such cases however, the deviation will be of such a character that the tracks are still acceptable. This will be so, for example, if the nuclear collision occurs near the mid-point of the trajectory of the particle in the air-gap. The observed deviation will then be less or greater than that due to the magnetic deflection alone, and the calculated values of the mass will be correspondingly greater or less than the true value.

Other processes, such as inelastic scattering of protons in the emulsion, will tend to lead to an over-estimate of the masses of the particles. In addition, as has been shown in §4, we must expect to observe a certain number of fortuitous coincidences, and their number in the particular conditions of the experiment should be 3 or 4. Because of these and other possible processes, it appears reasonable to attribute the observation of particles with intermediate values of the mass to scattering in the gas, or in the surface layers of the emulsion, or to other uncontrolled sources of error. The results are therefore regarded as giving no support for the view that other particles exist which are sufficiently stable to be brought to rest in solid materials, and which have masses in the interval between π-mesons and protons; and as proving that if such particles exist they either occur in matter bombarded by cosmic radiation at 11,000 feet, with a frequency less than 3 per cent of that of π and μ particles and less than 1 per cent of protons; or, alternatively, they have a lifetime shorter than 10^{-10} sec.

(η) Of the particles which originate in the emulsion (see fig. 5(a)), 11 have masses equal, within the limits of error, to that of protons. Three of the others may be identified as deuterons and one is of greater mass—probably an α-particle. The majority of the particles may be attributed to knock-on protons produced by the impact of neutrons with hydrogen nuclei; and others to nuclear disintegrations produced by neutrons of moderate energy, in which a single charged particle is emitted.

(ϑ) The distribution in the directions of motion of the μ^+ and μ^--mesons —see fig. 7 (a) and (b)—shows that most of these particles move downwards at a small inclination to the vertical —see (Camerini *et al.* 1948).

There is some indication that the maxima in the two distributions deviate from the vertical in opposite senses according to the sign of the charge. This may be accounted for if a considerable proportion of the particles are generated in the atmosphere or other material above the apparatus. They will then travel in the magnetic field for a distance of several centimetres before entering the emulsion, and will suffer a corresponding deflection.

The statistical weight of the angular distributions of the π^+ and π^--particles is too small for any conclusions to be drawn, but the observations are consistent with the assumption that the majority of particles are locally generated in the matter in the immediate neighbourhood of the plates, and that the directions of motion are distributed approximately isotropically.

TABLE V.

Type of particle	Number of particles	Number with			Percentage of all particles
		electron decay	slow electron	'blob'	
μ^+	30	19(24)*	—	—	$8{\cdot}5\pm1{\cdot}5$
μ^-	28 ± 4	4(18)*	6(18)*	6(18)*?	$8\pm2{\cdot}5$
π^+	14	—	—	—	$4{\cdot}0\pm1{\cdot}0$
π^-	19 ± 4	—	—	3	$5{\cdot}4\pm2{\cdot}0$
H^1	168	—	—	—	48 ± 4
H^2	55	—	—	—	16 ± 2
H^3,α etc.	39	—	—	—	11 ± 2

* These values are deduced from observations on electron-sensitive emulsions only. The figure in brackets gives the corresponding total numbers of particles observed in these plates.

The distributions for the protons, fig. 7(*e*), can be regarded as being made of two groups, one of which is due to particles moving in directions inclined to the vertical at less than 40°. From the displacement of the of the peak from the vertical, 10°, we must assume that most of the particles have travelled at least 5 cm. in the magnetic field. The other group, containing about 60 per cent of the total, is made up of particles of which the directions of motion are more widely dispersed, and they may be attributed to protons which originate in nuclear explosions produced in the matter of the apparatus.

The results of the present experiments are summarized in Table V., which shows details of the proportions of the particles of different types amongst all those on which measurements have been made.

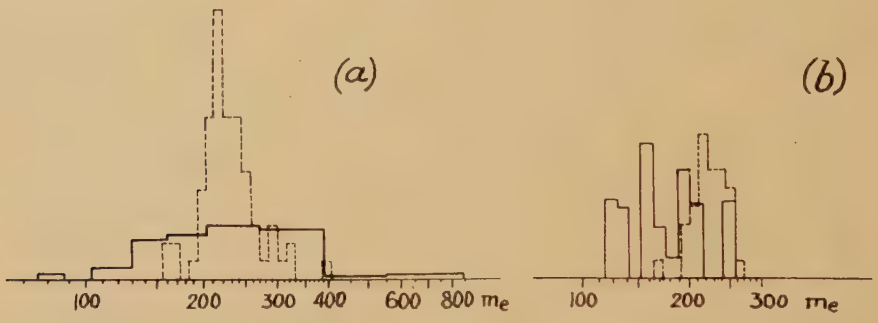
§6. Comparison with other Methods.

A comparison of the present results on the masses of the particles with those obtained in other experiments employing photographic

emulsions is shown in fig. 8. The figure gives the mass distributions obtained by Goldschmidt *et al.* from scattering measurements, and by Lattes, Occhialini and Powell from "grain-counts". In each case it can be seen that the "width" of the distribution, due to particles of a single type, is much greater than that obtained by the magnetic deflection experiments.

Experiments similar in principle to those described in the present paper have been carried out by Alichanian and Alichanov and their collaborators (1948). Using Geiger–Muller counters, they have measured the deflection, produced by a magnetic field, of particles brought to rest in a certain thickness of lead. The mass distribution thus obtained shows, (*a*) a prominent peak corresponding to the protons ; (*b*) groups of particles of greater mass than protons, but not individually resolved ;

Fig. 8.



Comparison with other methods. (*a*) The mass distribution of all the ρ-mesons, (*i. e.* μ^+, μ^- and that fraction of π^- which does not produce stars) obtained by magnetic deflection (dotted line), and that obtained by scattering measurements (full line) (Goldschmidt *et al.* 1948). (*b*) The mass distribution of the μ^+ obtained by magnetic deflection (dotted line), and that obtained by grain-counting (Lattes *et al.* 1948).

and (*c*) a continuous distribution towards lighter masses, which fills completely the interval between the groups due to protons and μ-mesons. The Soviet experimenters have interpreted their results as indicating the existence of many types of particles with intermediate masses which can carry both positive and negative charges. Further, a preliminary report has been given by Brode (1949) of experiments of the same type, but capable of greater precision, in which the deflections are measured by observations with Wilson chambers. In this case, it has been suggested, tentatively, that the observations may correspond to the existence of positive particles of mass about 1000 m_e. In the case of both these experiments the particles are arrested in layers of lead, and it is therefore not possible, in general, to observe any secondary processes occurring at the end of the range of the particles.

The conclusions to which we have been led by the results of the present experiments are therefore in contradiction with those obtained by other

methods. The seven observed particles with intermediate values of the mass have been attributed to uncontrolled sources or error. It appears to be of great significance that all but one of the particles are positively charged, and that they show no evidence of spontaneous decay at the end of their range. A similar feature is characteristic of the results described by Brode, in which all the observed particles of intermediate mass are positively charged. In view of the great variety of physical processes which can occur in apparatus of the type described by Alichanian and his colleagues, in addition to those which alone are assumed to be operative in the interpretation of the results, it may be suggested that the existence of other types of mesons, in addition to the π- and μ-particles, should be accepted with very great reserve; and that the present experiments prove that if they exist, they are certainly very much more rare than previous experiments have been taken to indicate.

Acknowledgments.

These experiments were originally undertaken in collaboration with Dr. S. Rosenblum, who installed the electro-magnet at the Jungfraujoch High-altitude Research Station, and who contributed both to the design of the apparatus and in making the exposures. I have great pleasure in acknowledging this substantial assistance without which it would have been impossible to complete the experiments.

I am greatly indebted to Professor C. F. Powell for his continuous help and encouragement both during the course of the work and in the discussion of the results. I also wish to thank Dr. H. Heitler for his help in the construction of some of the more delicate parts of the apparatus; Messrs. Dick and Roberts who assisted with the exposures and in the installation of the electromagnet; to Mrs. Andrews, Mrs. Reid and Miss Merritt for valuable assistance in the scanning of the plates and Professor Von Muralt and Mr. Widerkehr for kind hospitality. Finally, I wish to make grateful acknowledgment to Mr. Abele Giandolini of London, for financial assistance during the earlier part of the work.

References.

Alichanyan, Alichanov and Weissenberg, 1948, *Journ. Exp. Theor. Phys., U.S.S.R.*, **18**, 301.

Alichanyan, Alichanov, Morozov and Khrimian, 1948, *Acad. Sci. Rep.*, **61**, 35; and earlier references.

Bethe, 1946, *Phys. Rev.*, **70**, 821.

Bethe and Livingston, 1937, *Rev. Mod. Phys.*, **9**, 269.

Brode, 1949a, *Rev. Mod. Phys.*, **21**, 37; 1949b, *Communication at Como Conference, September.*

Camerini, Muirhead, Powell and Ritson, 1948, *Nature*, **162**, 433.

Goldschmidt-Clermont, King, Muirhead and Ritson, 1948, *Proc. Phys. Soc.*, **61**, 138.

Lattes, Occhialini and Powell, 1948, *Proc. Phys. Soc.*, **61**, 173.

Powell and Rosenblum, 1948, *Nature*, **161**, 473.

Retallack and Brode, 1949, *Phys. Rev.*, **75**, 1716.

Rossi and Greisen, *Rev. Mod. Phys.*, 1941, **13**, 240.

PLATES I. & II. (*a*).

Examples of " partners ".

Photographs of the tracks of particles at their exit from the first emulsion (first row), at the entry into the second emulsion (second row), and at the end of their range.

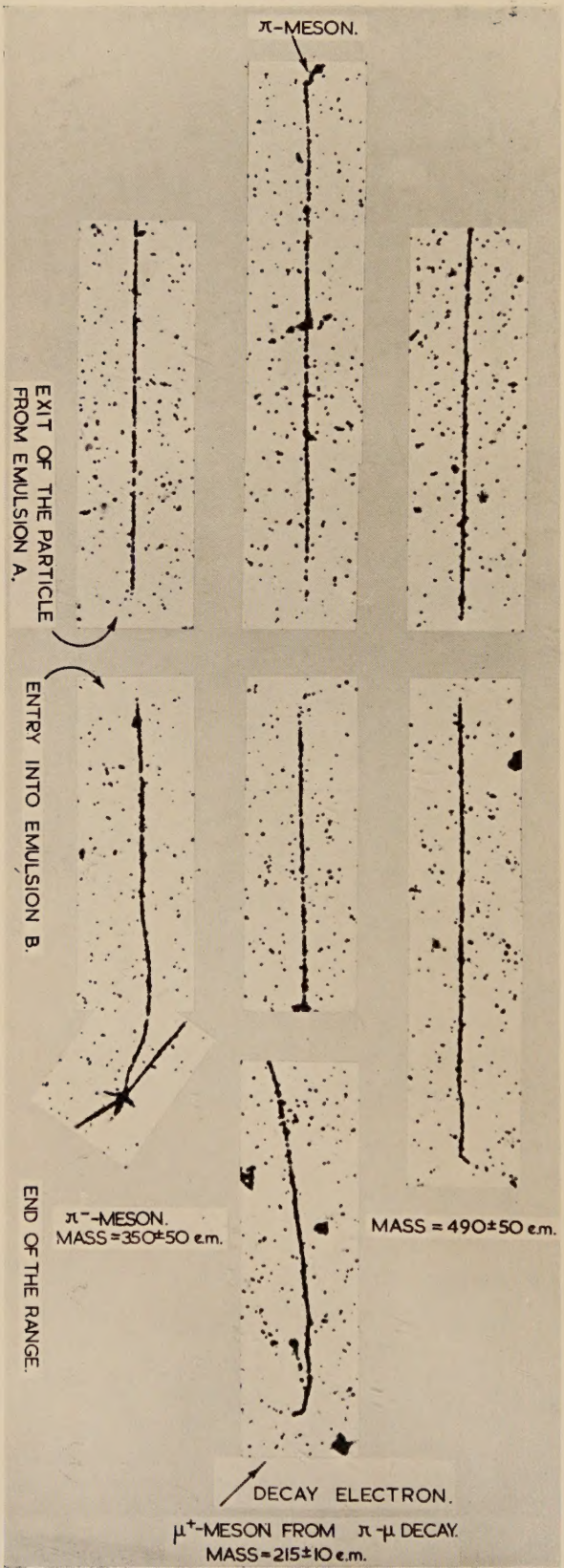
The great similarity between " partners ", which is immediately apparent under the microscope, has been partially lost in the photograph.

PLATE II. (*b*).

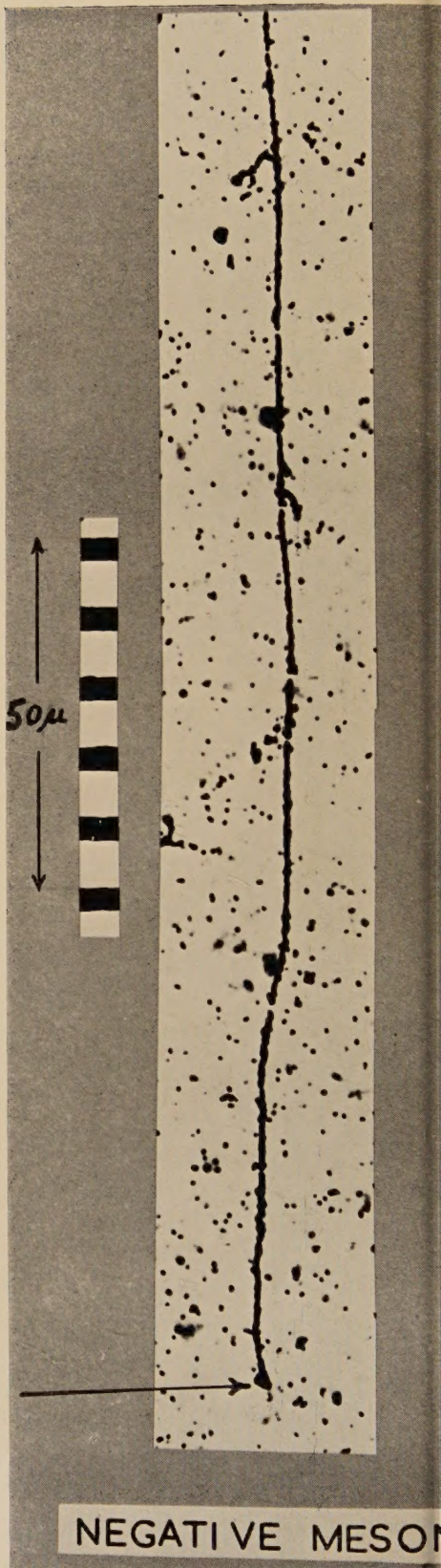
Negative meson. The arrow points to the " Blob " at the end of the range, which seems to be characteristic of negative mesons only.

To face page 106

(a)


π-MESON.
EXIT OF THE PARTICLE FROM EMULSION A.
ENTRY INTO EMULSION B.
π⁻-MESON.
MASS = 350±50 e.m.
MASS = 490±50 e.m.
END OF THE RANGE.
DECAY ELECTRON.
μ⁺-MESON FROM π-μ DECAY.
MASS = 215±10 e.m.

(b)


50μ
NEGATIVE MESO

VIII. *Notices of New Books and Periodicals received.*

Supersonic Flow Around Cones of Large Yaw. (Massachusetts Institute of Technology Technical Report No. 5.) [Pp. 125.] (Cambridge: Massachusetts.)

The properties of the flow about a yawing cone can be obtained as a power series in the yaw ϵ; earlier reports in this set (Nos. 1 and 3) presented tabular results for the coefficients of the terms independent of ϵ and linearly dependent on ϵ. The present report (in which the same value 1·405 for γ is used) deals with the coefficients of ϵ^2 for cones of semi-angle between 5° and 25° [5(2·5)15, 20, 25] and gives in Part I tables of the additional second order effects on velocity, pressure and density. Part II contains information for deriving drag (which has a second order contribution in ϵ), normal force, ratio of shock wave to cone yaw, and the coefficients determining the variations in entropy and in the shape of the shock wave which to this order ceases to be strictly circular in cross section. L. H.

The Origins of Modern Science. By Herbert Butterfield. (Bell & Sons.) Price 10*s*. 6*d*. net.

This work by the Professor of Modern History in the University of Cambridge is a welcome contribution to scientific literature.

A number of eminent scientists by their writings have demonstrated the cultural aspects of scientific thought and experiment. It is now gratifying to record the entry of a professional historian into this field in an endeavour to build a bridge between the humanities and the sciences.

In the reviewer's opinion the writer has achieved a conspicuous success. For this reason this is not the place to make minor criticisms. When the average student of science is so well versed in the history of his subject that he is capable of crossing swords with Professor Butterfield on his interpretation of some point in the views of, let us say, a sixteenth century physicist, then the bridge that the author is seeking to build will be truly fashioned. For this reason alone this volume is particularly to be commended to those who in Universities are concerned to see some broadening of the training of the specialized scientist.

An Advanced Treatise on Physical Chemistry.—Vol. I. *Fundamental Principles and Properties of Gases.* By J. R. Partington. [Pp. 943+40.] (Longmans.) Price 80*s*.

This book, as the author points out, is advanced in the sense of being comprehensive, and not in the sense of assuming much previous knowledge; for example, no previous knowledge of trigonometry or calculus is assumed, a concise account of these subjects being given in the mathematical introduction. This introduction is followed by sections on thermodynamics, kinetic theory of gases, statistical mechanics, quantum mechanics, thermometry, and a long final section on the properties of gases. Most sections start with a brief historical introduction and include a detailed account of the contributions of various workers in the field, with tables of experimental data and numerous references; the latter number about 18,000. Emphasis is laid throughout on the experimental side, though the theory is also thoroughly described. This book will be found a very valuable work of reference. A. F. D.

Applied Mathematics for Engineers and Scientists. By S. A. SCHELKUNOFF. [Pp. 472.] (D. van Nostrand Company Inc., 1948.) Price 36*s*.

PRESENT day science demands a mathematical discipline which often cannot be obtained from an undergraduate course. The author aims at providing this background for students of applied science. An idea of the ground covered can be obtained from the following list of contents: Complex numbers—Theory of approximation—Solution of equations—Power series—Differentiation—Integration—Vector analysis—Coordinate systems—Exponential functions—Differential equations of the first and higher orders—Partial differential equations—Conformal transformations—Contour integration—Linear analysis—The Gamma function—Exponential and Fresnel integrals—Bessel and Legendre functions—Formulation of equations. Thus the first half treats the general mathematical methods and the second half introduces the frequently occurring transcendental functions. The portion dealing with differential equations is particularly well written. The introductions to the use of Green's function and the Laplace transform method are two of the most valuable features of the book. The principles involved have been illustrated by concrete examples as far as possible. A chapter on tensors and matrices would perhaps have increased the general usefulness of the book. On the whole the book deserves to be widely used. H. C. B.

The Theory of Atomic Collisions. By N. F. MOTT and H. S. W. MASSEY. Second Edition, Royal 8vo. [Pp. 388.] (Oxford: The University Press, 1949.) Price 35*s*. net.

IN this work, which has been a classic since the publication of the first edition in 1933, the authors set themselves the task of discussing those collisions which do not involve creation or annihilation of particles, and can therefore be dealt with by wave-mechanics alone. In the second edition the form of the book remains unaltered, but the subject-matter is extended to include the phenomena of classical nuclear physics. There are valuable references to a large number of papers on both atomic and nuclear problems published since 1933, and the original treatment is expanded at varous points.

Broadly speaking, the book is divided into three parts. The first eight chapters are devoted to the general theory of collision processes, and include subsidiary mathematical and physical development where necessary. The main addition here is in the treatment of the dispersion formula, and of the collision complex or compound nucleus, although a more detailed analysis of the Born approximation has also been made.

Chapters IX. to XIII. deal with applications to atomic and nuclear collisions. More stress is now laid on nuclear problems and the chapter on the collision of electrons with molecules has been omitted.

In the last chapter relativistic two-body problems are considered. In orthodox quantum mechanics it has been usual to treat such problems by quantising the radiation field, thus bringing them outside the scope of this book, but the alternative method of the retarded potential has also been available, and the authors use this instead. Recently Feynman has criticised the use of field theory and developed the second method further, but it was obviously impossible to include a discussion of his work here. Some important work by Stueckelberg (1938), who foreshadowed Feynman's approach, has, however, apparently been overlooked by the authors. This chapter now includes a brief treatment of certain meson problems, together with useful references.

Altogether this new edition should prove invaluable to a wide class of physicists. K. V. R.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]